CRWR Online Report 07-01

Implementation of a State Hydrologic Information System

by

Tyler Landis Jantzen, M.S.

David R. Maidment, Ph.D.

May 2007

CENTER FOR RESEARCH IN WATER RESOURCES

Bureau of Engineering Research • The University of Texas at Austin
J.J. Pickle Research Campus • Austin, TX 78712-4497
This document is available online via World Wide Web at
http://www.ce.utexas.edu/centers/crwr/reports/online.html

Acknowledgements

I thank Dr. David Maidment for his guidance, support, and vision throughout this project.

I thank the Texas Natural Resources Information System for sponsoring this research, and for guiding my work. I especially thank Erika Boghici, Jim Scott and Chris Williams for answering numerous questions and helping me wade through hydrologic information in Texas.

May 2007

Abstract

Implementation of a Statewide Hydrologic Information System

Tyler Landis Jantzen, M.S.E.

The University of Texas at Austin, 2007

Supervisor: David R. Maidment

As part of its goal to "unite the nation's water information, to make it universally accessible and useful, and to provide access to the data sources," (*CUAHSI*, 2007) the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) has developed a Hydrologic Information System (HIS). An HIS is a tool that provides such uniform access to multiple sources of hydrologic data within a geospatial context, as described in CUAHSI's goal. While the CUAHSI HIS provides access to hydrologic data on a national scale, the need for access to statewide, regional and local hydrologic data has also been recognized. This thesis provides a background for hydrologic information technology, outlines the framework from which a statewide HIS should be created, and describes the Texas HIS prototype created in cooperation with the Texas Natural Resources Information System (TNRIS).

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AHPS	Advanced Hydrologic Prediction Service
CRWR	Center for Research in Water Resources
CSV	Comma Separated Values
CUAHSI	Consortium of Universities for the Advancement of Hydrologic Science, Inc.
dFIRM	Digital Flood Insurance Rate Map
DOQQ	Digital Orthophoto Quarter Quad
DRG	Digital Raster Graphic
EPA	Environmental Protection Agency
ESRI	Environmental Systems Research Institute, Inc.
ETL	Extract, Transform and Load
FEMA	Federal Emergency Management Agency
GEMStat	Global Environment Monitoring System
GIS	Geographic Information System
HIS	Hydrologic Information System
IMS	Internet Map Server
LCRA	Lower Colorado River Authority
LNRA	Lavaca-Navidad River Authority
MODIS	Moderate-resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NCDC	National Climatic Data Center
NAM12k	North American Mesoscale 12 Kilometer
NCEP	National Centers for Environmental Prediction

NED	National Elevation Dataset
NetCDF	Network Common Data Form
NEXRAD	Next-Generation Radar
NHD	National Hydrography Dataset
NHDH	High Resolution National Hydrography Dataset (1:24,000)
NHDPlus	National Hydrography Dataset Plus
NLCD	National Land Cover Data
NWIS	National Water Information System
NWS	National Weather Service
ODM	Observations Data Model
PLTS	Production Line Tool Set
SCAN	Soil Climate Analysis Network
SOAP	Simple Object Access Protocol
SQL	Structured Query Language
SRTM	Shuttle Radar Topography Mission
SSIS	SQL Server Integration Services
SSURGO	Soil Survey Geographic
STATSGO	State Soil Geographic
STORET	Storage and Retrieval
SWQM	Surface Water Quality Monitoring dataset
SWQMIS	Surface Water Quality Monitoring Information System
TAMUCC	Texas A&M University Corpus Christi
TCEQ	Texas Commission on Environmental Quality
TCOON	Texas Coastal Ocean Observation Network
TNRIS	Texas Natural Resource Information System

TRACS	TCEQ Regulatory Activities and Compliance System
TWDB	Texas Water Development Board
UN	United Nations
UNEP	United Nations Environment Programme
USGS	United States Geologic Survey
USDA	United States Department of Agriculture
WIID	Water Information Integration and Dissemination
WSDL	Web Service Definition Language

Chapter 1. Introduction

Hydrologic science, as it is practiced today, is largely dependent on measured hydrologic data. Information such as measured precipitation, evaporation, streamflow and groundwater levels describe the hydrologic cycle. Large amounts of data are required to better understand the complex interactions between different parts of this cycle. In essence, the more information that is gathered, the better the conclusions that can be synthesized, and a greater understanding of natural processes can be achieved.

With the recent technical developments in such areas as information technology, geographic information systems (GIS), relational databases and remote sensing, the amount of information available to hydrologists, and available through the internet, has exploded in the past two decades. However, the process of collecting and uniting these many types of data into single dataset and format that would provide the proper framework for further analysis and synthesis is tedious, time consuming, and often the major barrier to the advancement of hydrologic science. In fact, when a group of hydrologists and other hydrologic data users were polled, 36 percent said that they spend more than 25 percent of their time acquiring and preparing data (*Maidment*, 2005). Collecting data from multiple sources often means going to multiple websites, creating multiple user accounts, and requesting data in multiple formats. These data come in different projections, for different date ranges, at different spatial and temporal resolutions, and for different geographic areas. It is easy to recognize that combining these data into a single workspace is not a simple task.

The development of a Hydrologic Information System (HIS) is a solution to many of the data gathering problems that currently exist with hydrologic data. An HIS is a generic term used to describe a hydrologic information access tool. The HIS being described by the majority of this document is a map-based web portal providing access to numerous types of hydrologic information in a consistent, easy to use format. The same term has taken on similar but slightly different meanings when used by other projects such as the HIS created by The University or Arizona (*SAHRA*, 2006) and that created by the University of Nebraska (*Soh et al.*, 2006).

The HIS provides the method and framework for comprehensive searching and accessing of hydrologic data within a geospatial context. It brings together data of different types (e.g. precipitation and streamflow) that are presently stored in separate, non-communicating relational databases, and thematically integrates this data across time, space, and inter-agency boundaries (see Figure 1.1). It also allows for data of the same type (e.g. precipitation) stored by different entities to be combined and integrated into a more complete hydrologic picture. These capabilities are currently unavailable elsewhere. While some tools provide access to multiple sources of data, none provides the framework for such comprehensive data integration as the HIS.

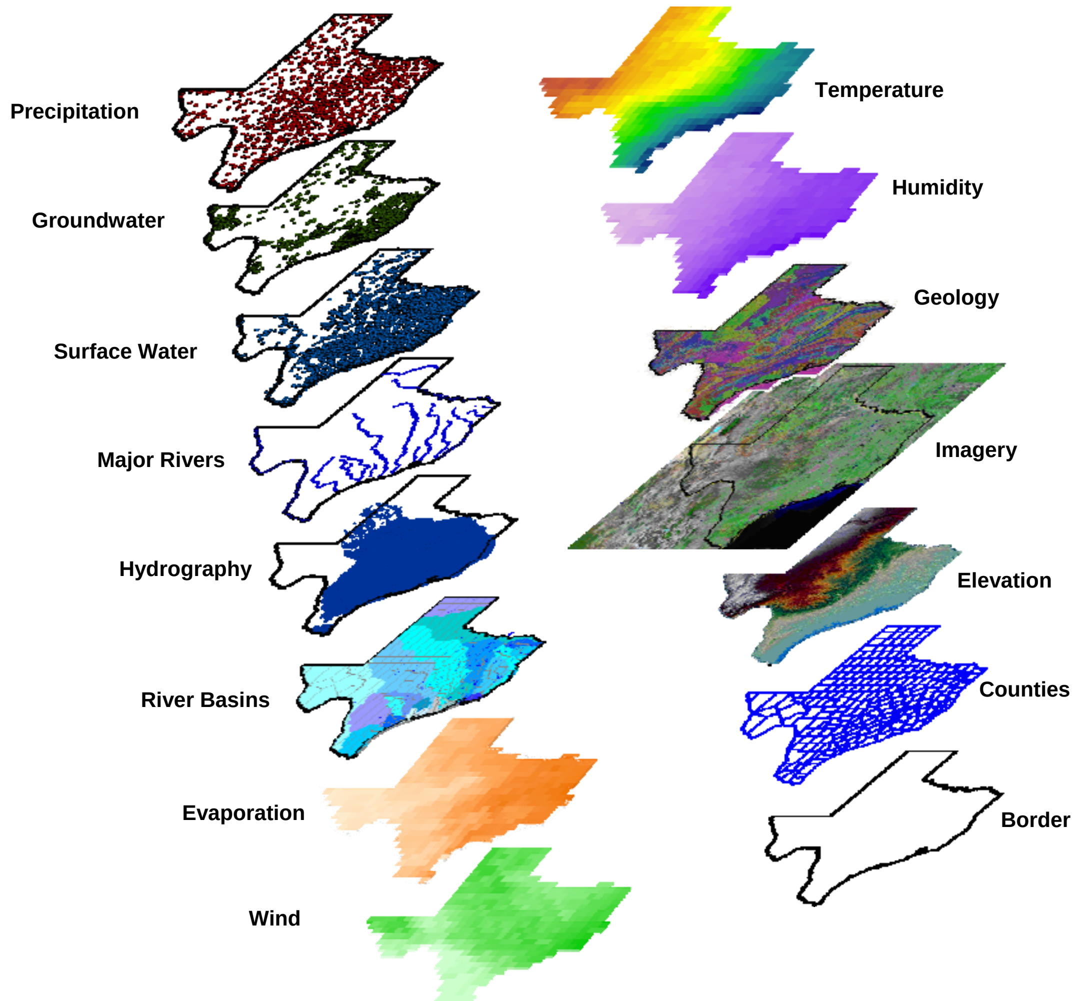


Figure 1.1 Data Layers in the Texas HIS

These concepts are demonstrated by the prototype created for the state of Texas, using surface water quality monitoring data and streamflow lines. Given a common format and method of access for multiple sources and types of hydrologic information, common analysis tools can be developed. The goal of an HIS is to reduce the time and energy spent collecting and uniting hydrologic information, and to increase the amount and variety of information readily available for applied use. With such an increase in data efficiency, analysis and synthesis of new hydrologic science concepts can become a greater focus leading to a better understanding of hydrologic phenomena. The following document describes the overall concept of the HIS, as well as the design of a prototype statewide HIS for the use by the state of Texas.

1.1. USES OF HYDROLOGIC INFORMATION

Hydrologic information can be defined as information or data that describes or pertains to the hydrologic cycle (see Figure 1.2). This includes, but is not limited to, data describing precipitation, evapotranspiration, energy flux, aquifers, groundwater infiltration, wells, surface flow and storage, stream networks, land use and land cover, topography, watersheds, hydraulic structures, climate and water quality. This information is essential in the development of important hydrologic models and data products such as floodplain maps, drought and flood forecasts, water rights determination, and comprehensive municipal plans.

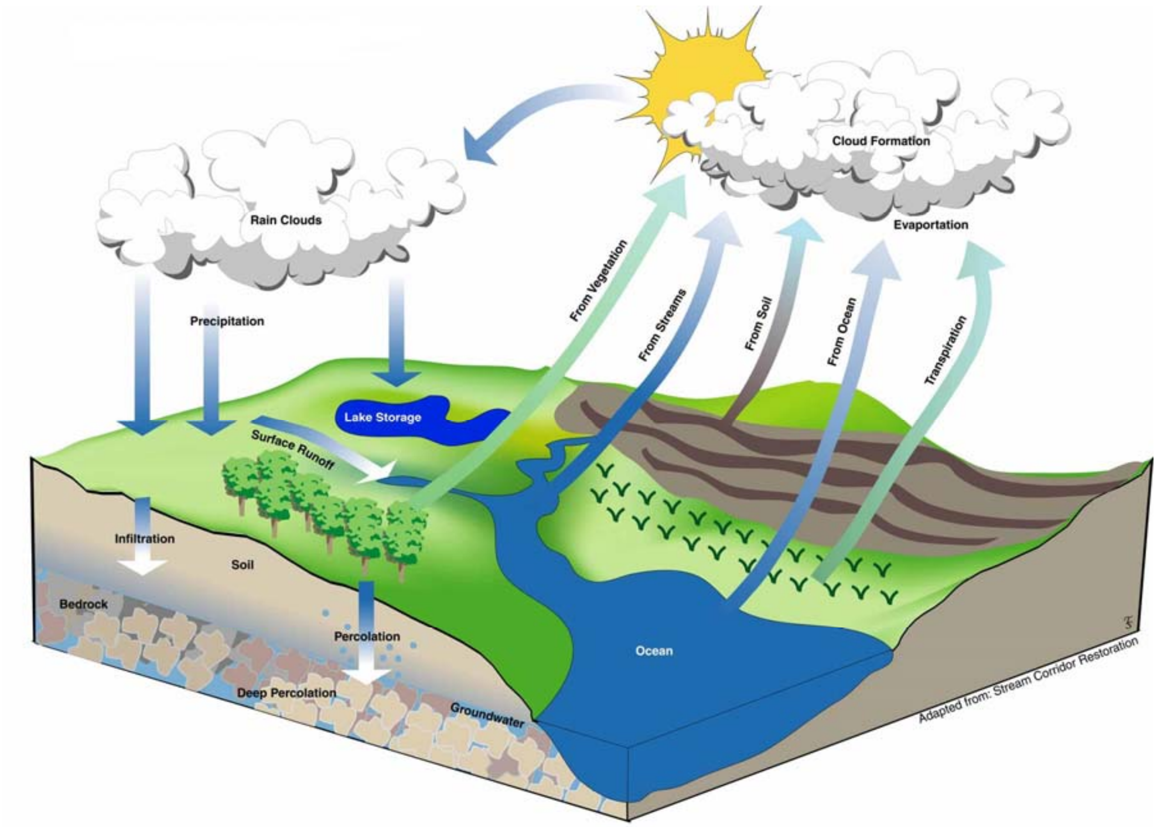


Figure 1.2 Hydrologic Cycle (*Iowa State University*, 2007)

Hydrologic data is needed for a wide range of applications. Some users need only a single value, or one type of information. For instance, a concerned property owner may want to know the historic and present level (and thus streamflow rate) of a nearby stream. An irrigation farmer may want to know the groundwater level at wells near his farm. A regional water planner may want to know the historical rainfall for a given area.

Some users need multiple types of information. A land development engineer may want both historical rainfall, streamflow and land use data to determine the possible effects of new real estate. A hydraulic engineer for a levee board may need rainfall and streamflow data for the entire upstream watershed, as well as probability statistics for future events. A hydrologic researcher may need dozens of types of information as inputs into a complex hydrologic model.

Hydrologic information is needed by a wide variety of users. People working in academia, government, and private business all have hydrologic information needs. Private citizens also have hydrologic information needs. While the type of information needed may be the same, the use, level of experience, and amount of information are different for each user.

For an HIS to be useful, it must meet the needs of various types of users. The HIS prototype discussed in this document was designed primarily with the academic and governmental users in mind. However, as additional data types are included, the needs of other types of users will also be addressed.

1.2. Types of Hydrologic Information

Hydrologic information can typically be categorized under one of three types: geospatial information, time series information, and space-time composite information. Geospatial information refers to that which is specifically "projected across the earth's surface and distributed in space" (*Ruddell and Kumar*, 2006). An example is most any kind of information that appears on a map. Time series information records that which is distributed across time, such as change in global population or fluctuations in stream flow. The integration of these two types of data into space-time composite datasets has typically created numerous difficulties, both in modeling and understanding. However, certain remote sensing products such as the NEXRAD radar animation seen on the evening news have successfully integrated both space and time. Visualizing four dimensions (three spatial dimensions and one temporal dimension) or more is challenging (*Goodall*, 2005). For instance, one can easily visualize the temporal variation of precipitation at a single point in space (see Figure 1.3). One can also easily visualize the two dimensional spatial variation of precipitation at a single point in time (see Figure 1.4). Modern computerized animation technology has made the visualization of three and four dimensions a reality. The graphic in Figure 1.4 can be modified to be animated through time, with each frame representing a different period. Creating an animated representation of three spatial dimensions has also been accomplished, but must be carefully adjusted to be readily understood.

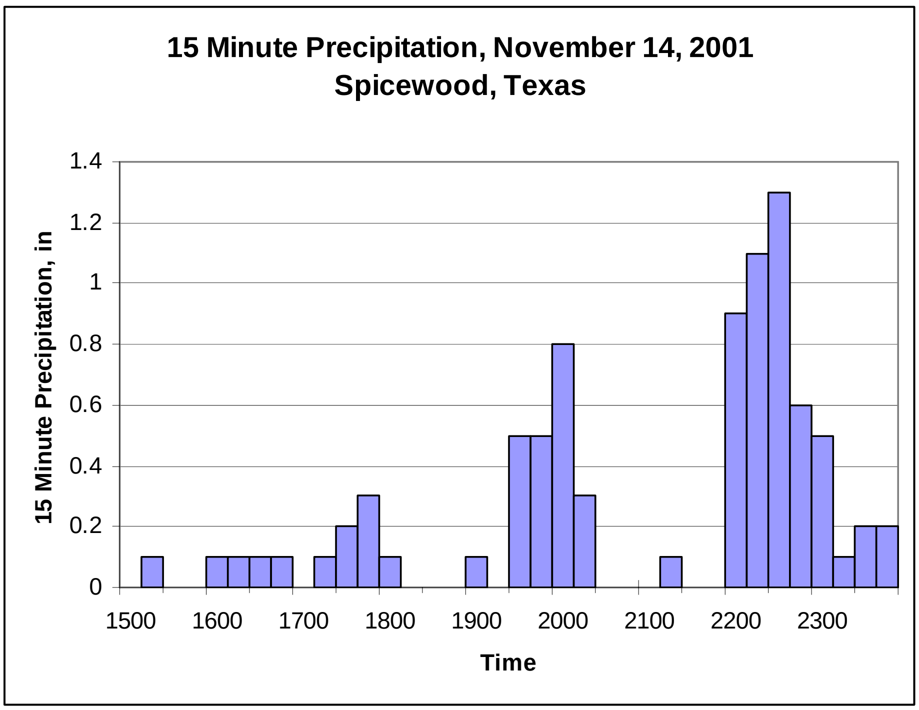


Figure 1.3 Precipitation temporal variation at a single point. Data from *NOAA National Data Center*, 2007

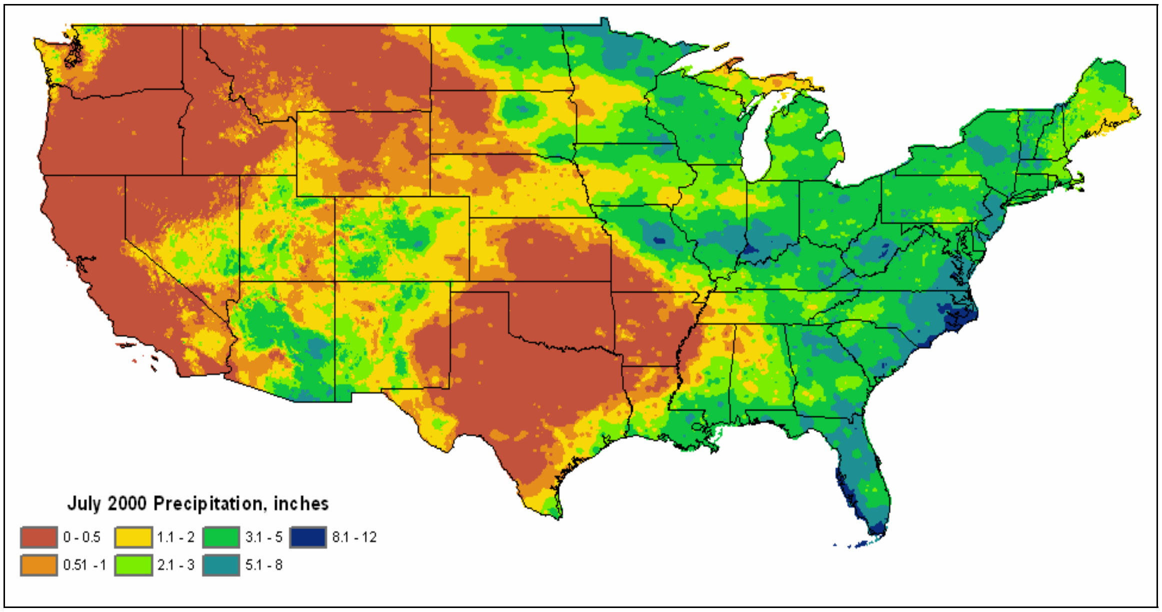


Figure 1.4 Two spatial dimensions at a single point in time. Data from *PRISM Group*, 2006

Similar to visualizing, difficulty exists in using three and four dimensional data in computerized models. Advances in multi-dimensional array data storage, such as the netCDF format, now allow such data to be used in models. Software that fully utilizes these data formats is still being developed. It is anticipated that significant advances in multi-dimensional data visualization and modeling will be made in the near future.

While it is important to understand the implications of working with three and four dimensional data, this document, and the current prototype for a Hydrologic Information System will focus on using two different types of two-dimensional information: time series data at a single point in space (using the Observations Data Model), and static geospatial data (using polygon and raster fields in NHDPlus).

1.3. Sources of Hydrologic Information

The relatively recent rise in information sharing technology has not bypassed the field of Hydrology. Enormous amounts of information are now available through the internet. This information is distributed across hundreds of individual sources. Any attempt to review all such sources would be incomplete. However, these sources can be categorized, and each category qualitatively described. The following is a discussion of sources of hydrologic information, specifically as it applies to the creation of a statewide HIS.

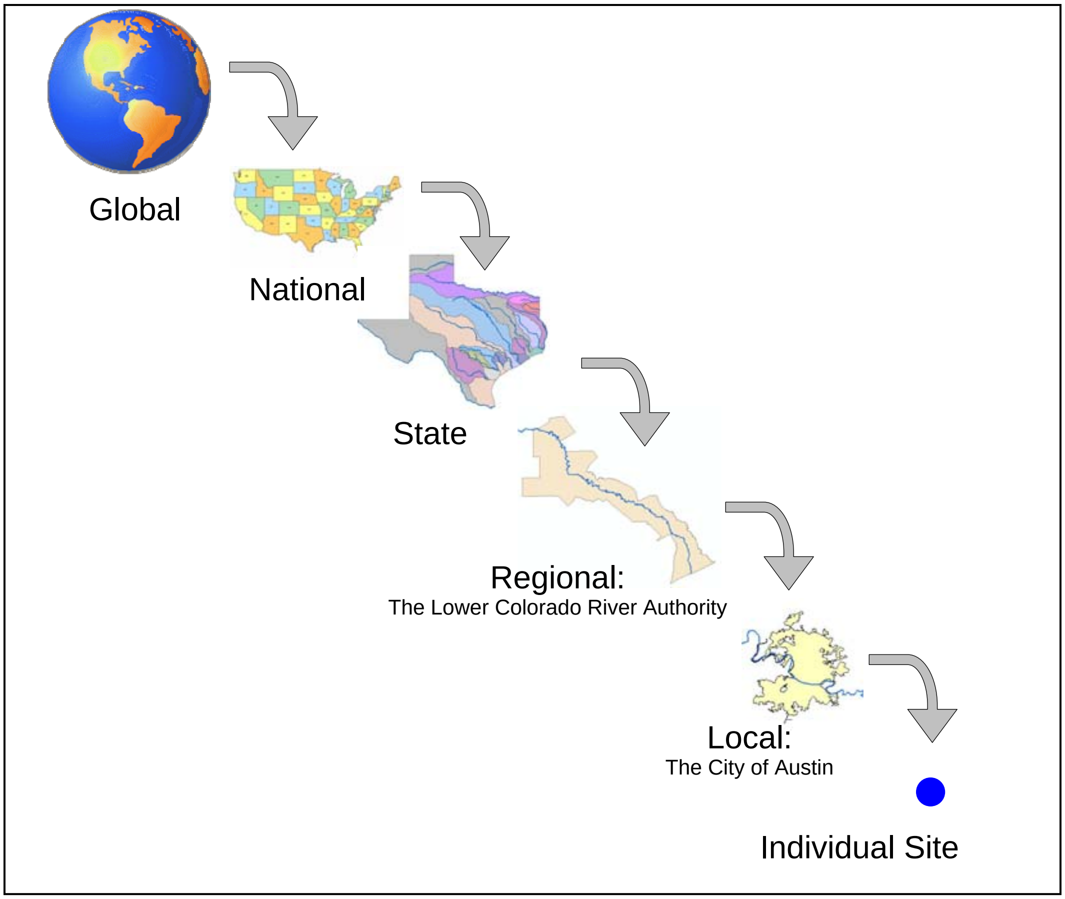


Figure 1.5 Spatial Scales of Hydrologic Information

Hydrologic information is collected at five key spatial scales: global, national, state, regional and local (see Figure 1.5). Information at each of these scales is collected, managed, maintained, and funded by inherently different types of groups. Access to these different types of sources is currently scattered. The goal of an HIS is to begin to integrate these disparate sources. The differences among these different types of sources have wide implications for the integration of such sources through an HIS.

1.3.1. Global

Various sets of hydrologic information are available at a Global or near-global scale. Due to the large spatial scale, these sources are limited to those collected by remote sensors. Access to such information is provided by such groups as the United Nations (UN) and the National Aeronautics and Space Administration (NASA). This information is typically provided at a coarser resolution (1 km grid or larger) than smaller datasets. A brief listing of some of the global hydrology data sources is provided in Table 1.1.

Table 1.1 Global Hydrology Data Sources

Agency	Dataset	Type of Data	URL
UNEP	GEMStat	Point observations of water quality and streamflow across the globe	http://www.gemstat.org/
NASA	MODIS	Geospatial atmospheric, land and oceanic data from satellite sensing	http://modis.gsfc.nasa.gov/

1.3.2. National

Many countries have national programs for the collection of hydrologic data. This is especially true within the United States, where a plethora of national agencies collect hydrologic information. The scopes and missions of these sources will not be discussed here. Instead, a brief list of agencies and types of information is provided in Table 1.2. Some of these data sources are in the process of being provided via a national HIS. The integration of national and state HIS will be discussed in section 3.9. Depending on the availability of state, local and regional sources, national sources may be considered for inclusion in a statewide HIS.

Table 1.2 National Hydrology Data Sources

Agency	Dataset	Type of Data	URL
USGS	NHD	Geospatial surface water information	http://nhd.usgs.gov/
USGS	NWIS	Point observations of streamflow, groundwater and water quality	http://waterdata.usgs.gov/nwis
EPA	STORET	Point observations of water quality, biological and physical data	http://www.epa.gov/storet/
EPA	NHDPlus	Geospatial surface water, land use and elevation information.	http://www.horizon-systems.com/nhdplus/
NCDC		Numerous weather and climate related datasets	http://www.ncdc.noaa.gov/oa/ncdc.html
NWS	AHPS	Point observations of historical and forecasted (+24 hrs) stream flow	http://www.nws.noaa.gov/oh/hic/
NASA/ USGS	SRTM	Geospatial topography data for the world.	http://srtm.usgs.gov/
USDA	PRISM	Geospatial precipitation, temperature, and other climate data	http://www.wcc.nrcs.usda.gov/climate/prism.html
USDA	SCAN	Point time series soil water balance data	http://www.wcc.nrcs.usda.gov/scan/
USDA	SNOTEL	Point time series snowpack data	http://www.wcc.nrcs.usda.gov/snow/
USDA	SSURGO, STATSGO, WBD etc.	Numerous soil and watershed related geospatial datasets.	http://www.ncgc.nrcs.usda.gov/products/datasets/index.html
CUAHSI	HIS	National HIS Portal	http://www.cuahsi.org/his/
FEMA	dFIRM	Geospatial floodplain information viewer	https://hazards.fema.gov/femaportal/wps/portal
Dept. of Energy	AmeriFlux	Point observations of water, heat and carbon flux in the atmosphere	http://public.ornl.gov/ameriflux/
NCEP	NAM12k	Modeled forecasted climate grid data.	http://motherlode.ucar.edu:8080/thredds/catalog/model/NCEP/NAM/CONUS_12km/latest.html

1.3.3. State

Most states have data collection efforts independent of the national data collection structure. Depending on many bureaucratic factors, these data may or may not be integrated with a corresponding national data source. Most details of statewide data collection, including funding, extent, type of data and public availability, vary on a state-by-state basis. A list of potential sources for inclusion in the example statewide Texas HIS is included in Table 1.3.

Table 1.3 Texas Statewide Hydrologic Data Sources

Agency	Dataset	Type of Data	URL
TNRIS		Numerous geospatial datasets, including imagery, hydrography, geology and political boundaries	http://www.tnris.state.tx.us/datadownload/download.jsp
TCEQ	SWQM	Point observations of water quality data	http://www.tceq.state.tx.us/compliance/monitoring/crp/data/samplequery.html
TCEQ		Geospatial surface water information: Atlas of Texas Surface Waters	http://www.tceq.state.tx.us/implementation/water/tmdl/atlas.html
TCEQ		Geospatial surface water information: Surface Water Quality Viewer	http://www.tceq.state.tx.us/compliance/monitoring/water/quality/data/wqm/viewer/viewer.html
TWDB		Numerous geospatial datasets, including aquifers, well locations,	http://www.twdb.state.tx.us/mapping/gisdata.asp
TWDB	WIID	Surface and groundwater point data viewer	http://wiid.twdb.state.tx.us/
TWDB		Monthly evaporation and precipitation data by quadrangle	http://hyper20.twdb.state.tx.us/Evaporation/evap.html
TAMUCC	TCOON	Point observations of coastal hydrology, including salinity, water level, and climate	http://lighthouse.tamucc.edu/pq

1.3.4. Regional

Many states are sub-divided into smaller regions. These smaller regions collect, manage, and provide hydrologic data at a regional level. These subdivisions include water resource management districts such as the Lower Colorado River Authority (LCRA) in Texas, and terrain-defined hydrologic regions. Like the statewide sources, this varies on a state-by-state basis. A list of a few relevant regional sources in Texas is included in Table 1.4. These sources, and others of similar scale, should be considered for inclusion in a statewide HIS.

Table 1.4 Texas Regional Hydrologic Data Sources

Agency	Dataset	Type of Data	URL
LCRA	HydroMet	Point observations of streamflow, precipitation, humidity and temperature	http://hydromet.lcra.org/index2.shtml
Regional Water Districts		Geospatial information about the Texas Regional Water Districts	http://www.texaswatermatters.org/regions.htm
Brazos River Authority		Reservoir capacity, storage, and release information	http://www.brazos.org/waterSupply.asp
LNRA		Reservoir capacity, storage, and release information	http://www.lnra.org/reservoirdata.asp
Texas Water Info		Links to various river authorities, including surface water quality and flow data.	http://www.texaswaterinfo.net/Monitoring/SW/RA%27s.htm

1.3.5. Local

The smallest level of hydrologic information scale is the local scale. This includes county, city, parish, township and other similar local jurisdictions. It also includes individual site data collected by principal investigators, scientists, and research projects. These groups often collect, manage, and provide access to various types of hydrologic information. A comprehensive list of the hundreds of possible local data sources in Texas is not provided in this document. However, a list of a few such sources for consideration in a statewide HIS has been provided in Table 1.5.

Table 1.5 Texas Local Hydrologic Data Sources

Agency	Dataset	Type of Data	URL
City of Austin		Point observations of water quality	http://www.ci.austin.tx.us/wrequery/db_query_form.cfm
Harris County Flood Control District		Watershed boundary data	http://www.hcfcd.org/webprogram.html
Paul Montagna		Point observations of water quality	Personal contact

1.3.6. HIS Data Integration Models

The goal of a statewide HIS is the integration of hydrologic data sources from within each of these spatial scale levels, and also among organizations of different scales. Various models exist for this integration.

One model provides for the hierarchical information transfer between spatial levels. In this model, each successively higher level is responsible for the integration of relevant data from the next lowest level. This is shown in Figure 1.6. One advantage of a hierarchical model is that each coordinating agency only has to integrate data from a small number of sources from the level directly below. Disadvantages to such a model include the creation of numerous small HIS's, and the tendency to lack common unifying data standards.

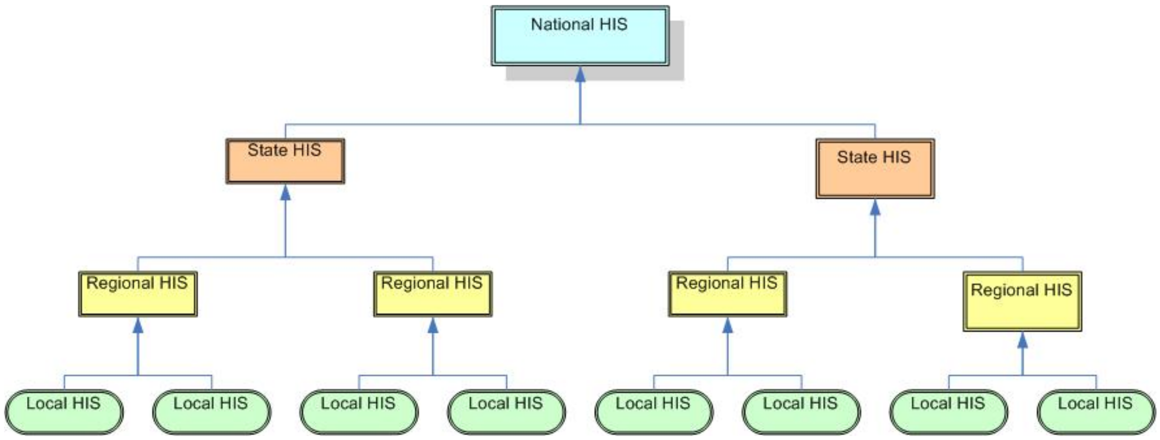


Figure 1.6 Hierarchical HIS Model

Another model is the distributed federated model, whereby all distributed sources from all levels are integrated (or federated) by a single agency's HIS. This is shown in Figure 1.7. An advantage of such system is that a single data collection and provision standard would be easy to implement. This is the model under which the national HIS, coordinated by CUAHSI, is currently operating. However, a single agency could potentially be responsible for collecting and integrating data from hundreds of disparate sources, a task that would definitely be overwhelming, tedious and potentially impossible. For instance, if the Texas HIS operated under a distributed federated model, it may potentially be responsible for integrating data from 254 counties, 16 groundwater management districts, 16 regional water planning areas, and numerous large cities and municipalities. In practice, some combination of the hierarchical and distributed federal model may need to be implemented, ensuring both continuity of standards and reasonable levels of responsibility.

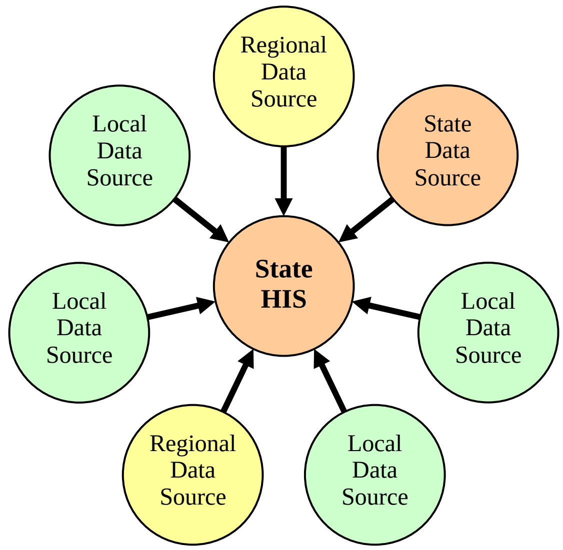


Figure 1.7 Distributed Federated HIS Model

Most sources already incorporate some degree of inter-level data integration. For instance, the Texas Commission on Environmental Quality (TCEQ) Regulatory Activities and Compliance System (TRACS) database includes the Surface Water Quality Monitoring (SWQM) dataset collected both by the multiple sampling programs within the TCEQ, and also by other collaborating partners from local and regional levels (*Texas Commission on Environmental Quality*, 2006). Thus, SWQM provides a statewide, comprehensive water quality time series dataset. However, in the context of a statewide HIS, the issue arises as to how to interpret those data along with other state data sources.

Despite some degree of inter-level data integration, there remain many sources of similar data type from different spatial levels that are not integrated, and thus provide an incomplete data picture. The United States Environmental Protection Agency (EPA) Storage and Retrieval (STORET) program collects and provides access to water quality information across the country, similar to how TRACS integrates statewide water quality in Texas. Many of the actual sampling sites within the STORET database are owned and operated by organizations other than the EPA. Figure 1.8 shows a map of all the STORET sites across the continental U.S. based on the STORET sites records in July, 2006. Notice that while many states seem to have very dense data coverage, some states such as Texas, Mississippi, Alabama and Virginia have very scarce data coverage. This is not because water quality data is not measured in these states. Instead, the existing water quality information is not sufficiently integrated across spatial levels.

Figure 1.9 shows a map of the TRACS SWQM sampling locations within Texas. This map is the equivalent of the previous figure of EPA STORET sampling sites, but on a state level. In the case of Texas, water quality data collection began in 1967, before the creation of the EPA in 1970, and has been maintained ever since. Efforts are currently under way to modernize both the TRACS and STORET systems from within each respective agency, to facilitate better data integration. The modernized version of the TRACS surface water quality system is called Surface Water Quality Monitoring Information System (SWQMIS). The modernized version of STORET is called Water Quality Exchange (WQX), (*USEPA*, 2007). While the EPA and TCEQ work to integrate their systems, another method of integration can be provided by an HIS. Figure 1.10 shows a map of sampling sites, showing what integration provided by an HIS might look like.

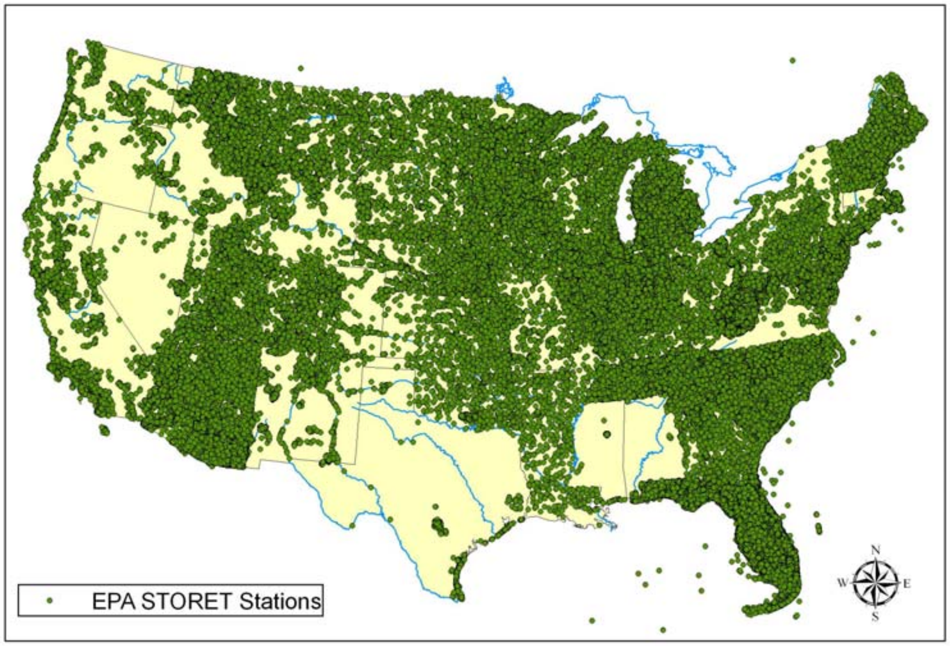


Figure 1.8 EPA STORET Locations.

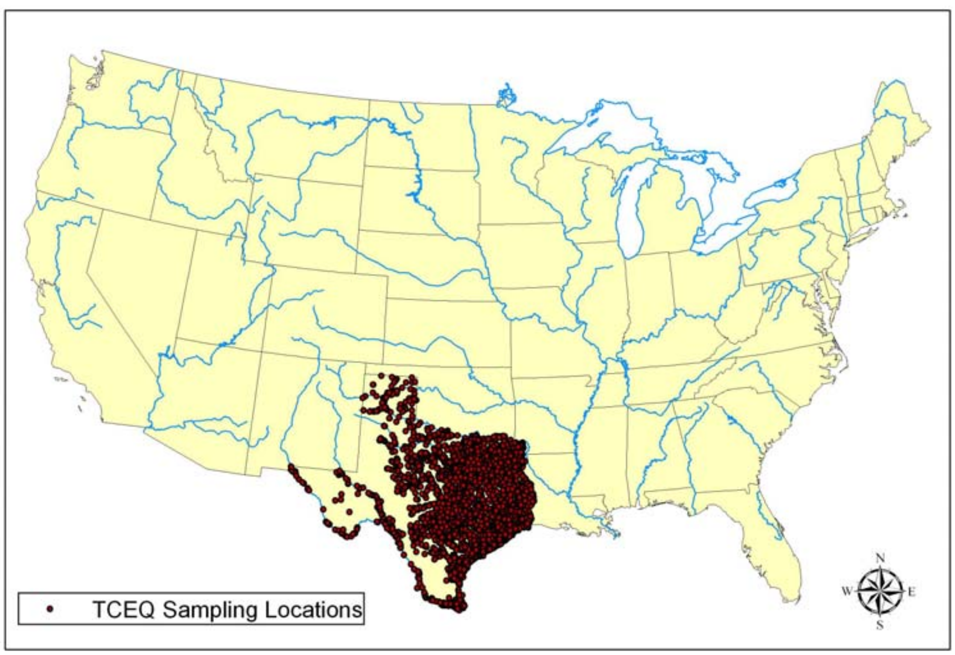


Figure 1.9 TCEQ SWQM Sampling Locations

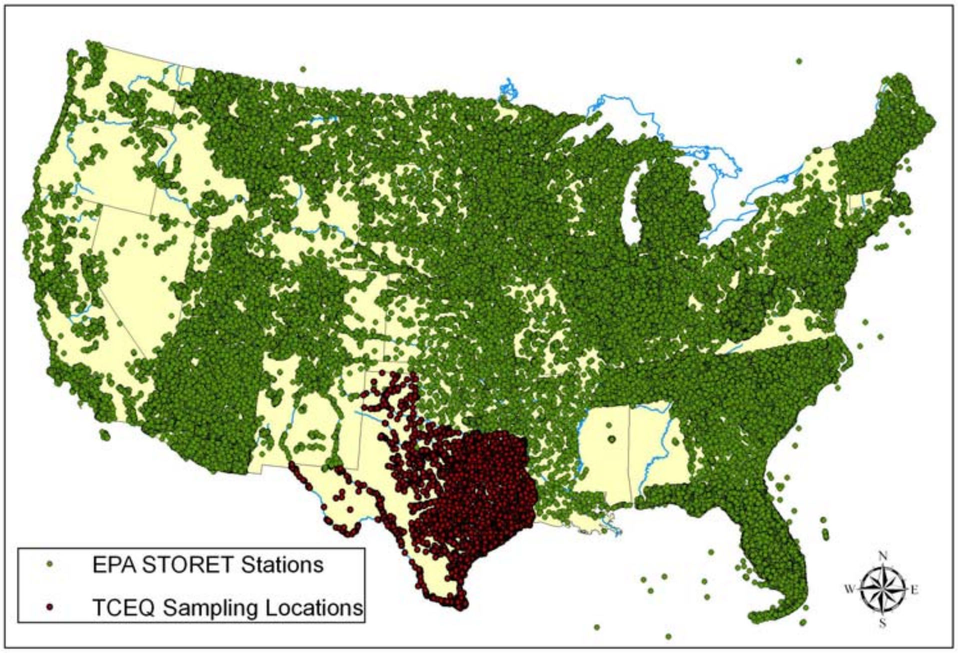


Figure 1.10 EPA STORET and TCEQ SWQM Locations

1.4. FORMATS OF HYDROLOGIC INFORMATION

As hydrologic information is collected by a wide range of groups and agencies from multiple spatial levels, so is it also collected and stored in a wide variety of formats. These formats are often incompatible with each other. One of the main challenges to hydrologic science is the reconciliation of these incompatible formats. This is primarily a software challenge, but has its roots in the manner in which data is stored and processed, as well as the inherent data type. This document will not attempt to review in any completeness the theory behind data storage or interoperability. Such a review would quickly become obsolete due to rapid advances in software and technology, and is beyond the scope of this project. A brief list of hydrologic data types that may be encountered in the development of a statewide HIS is included in Table 1.6. Additional background to hydrologic data formats and interoperability can be found in Chapter 8 of *Folk*, 2006.

Table 1.6 Hydrologic Data Formats

Type	File Extension	Proprietary Software	URL
Comma Separated Value	.csv	N/A	http://tools.ietf.org/html/rfc4180
Microsoft Excel Spreadsheet	.xls	Microsoft Excel	http://office.microsoft.com/excel
Microsoft Database	.mdb	Microsoft Access	http://office.microsoft.com/access
SQL Server	.mdf	Microsoft SQL Server	http://www.microsoft.com/sql/
ASCII / Data	.dat	N/A	
ASCII / Text	.txt	N/A	
Tab delimited	N/A	N/A	
Other delimited	N/A	N/A	
Shapefile	.shp	ESRI	http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf
Raster	.jpg, .tif, .gif, etc.	N/A	
NetCDF	.nc	N/A	http://www.unidata.ucar.edu/software/netcdf/docs/faq.html
File Geodatabase	.gdb	ESRI	http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=Types_of_geodatabases
Extensible Markup Language	.xml	N/A	http://www.w3.org/XML/
Spatial Data Transfer Standard	.sdts	N/A	http://mcmcweb.er.usgs.gov/sdts/whatsdts.html

Increasingly, more and more of these formats are becoming interoperable. However, even if interoperable, transferring data from one format to another is often tedious and labor intensive. Such a transfer usually requires an in depth understanding not only of the format, but also of the contents of the data. An example of data format conversion is included in the description of the integration of the SWQM dataset into the Texas HIS discussed in section 4.4.

1.5. OBJECTIVES

The objective of this thesis is to answer the following questions:

- What is the current state of hydrologic data and technology as it applies the creation of an HIS?
- What does a framework for the creation of a statewide HIS that provides hydrologic information in a consistent and easy to use format look like?
- How can this framework be applied to an HIS for the state of Texas?
- What is required to build a prototype of this system?
- How should data be added to this system?
- How can this system be integrated with a national HIS?
- What direction should further research take?

Chapter 2. Technology and Literature Review

The creation of a statewide HIS is intrinsically linked to modern information, database, and modeling technology. The implementation of the system described in this thesis is dependent on such technologies as databases, web services and GIS. The dynamic nature of these technologies is recognized. However, it is essential to understand the current technological elements related to an HIS before fully understanding the HIS itself. This section briefly reviews such technologies in an attempt to provide context for the creation of an HIS. Additionally, no new concept exists in a bubble. This section also provides a brief review of the intellectual context for an HIS.

2.1. OBSERVATIONS DATA MODEL

One of the key elements in information science is the data or database model. A database model is a "standardized structure that organizes data" (*Ruddell and Kumar*, 2006). One of the key developments in hydrologic information science (or hydroinformatics) has been the recent development of the Observations Data Model (ODM). The ODM was developed as part of the CUAHSI national HIS project (*Tarboton et al.*, 2006). The ODM has been designed as a relational database schema in which to store point observation time series data in a meaningful way. The ODM is designed to maintain not only the observations data, but also significant supplementary information about the data, otherwise known as metadata. Figure 2.1 shows the general schematic of the ODM release version 1.0 The key tables (DataValues, Variables and Sites) are surrounded by multiple auxiliary tables with supporting information.

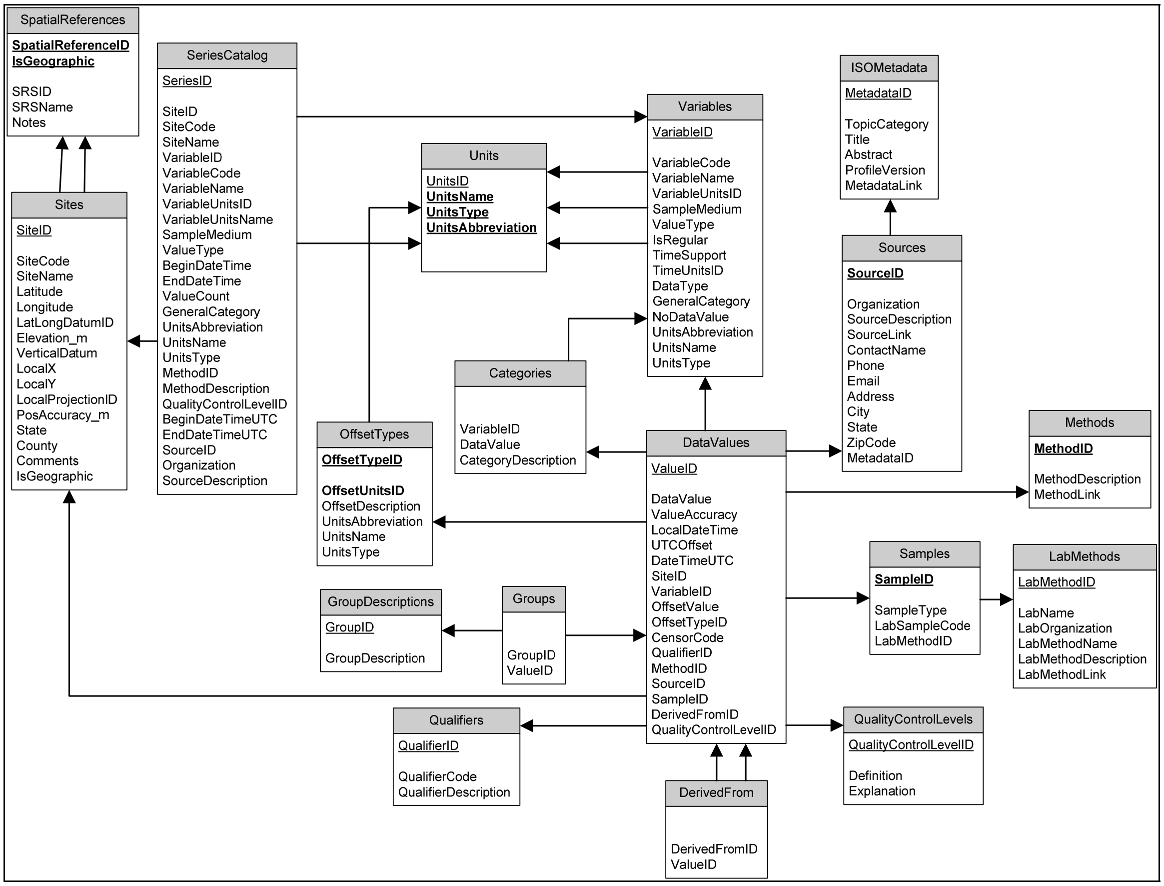


Figure 2.1 Observations Data Model Version 1.0 Schema

The use of a common data model has multiple advantages within the context of a spatially distributed information network, such as a national or statewide HIS. If all information is stored and maintained in a standard format, access to such information also becomes standard. As new applications and tools are built for analyzing data in this standard format, all data that is held in this format can be mobilized by the same application.

An example of differences in data models occurs in the analysis and plotting of stream gage and precipitation measurements. A stream gage and a rain gage measure inherently different types of information. Stream flow is measured as instantaneous values, often measured on a regular period and averaged over time. Most USGS streamflow gages measure streamflow every 15 minutes to then take daily averages of these values to produce daily streamflow. Precipitation gages measure the amount of rain that has fallen since the last measurement. These measurements may happen on a regular schedule, but more often are made only after a measurable amount of precipitation has been collected, i.e. only after it has rained.

Because these observations are measured at different time scales, and typically at different sites, the data itself is typically stored in different formats and data models. However, if it were all to be stored in a standard data model, precipitation and streamflow could be easily accessed and compared in a single application, leading to easier and faster analysis of the transformation between the precipitation input and the streamflow output.

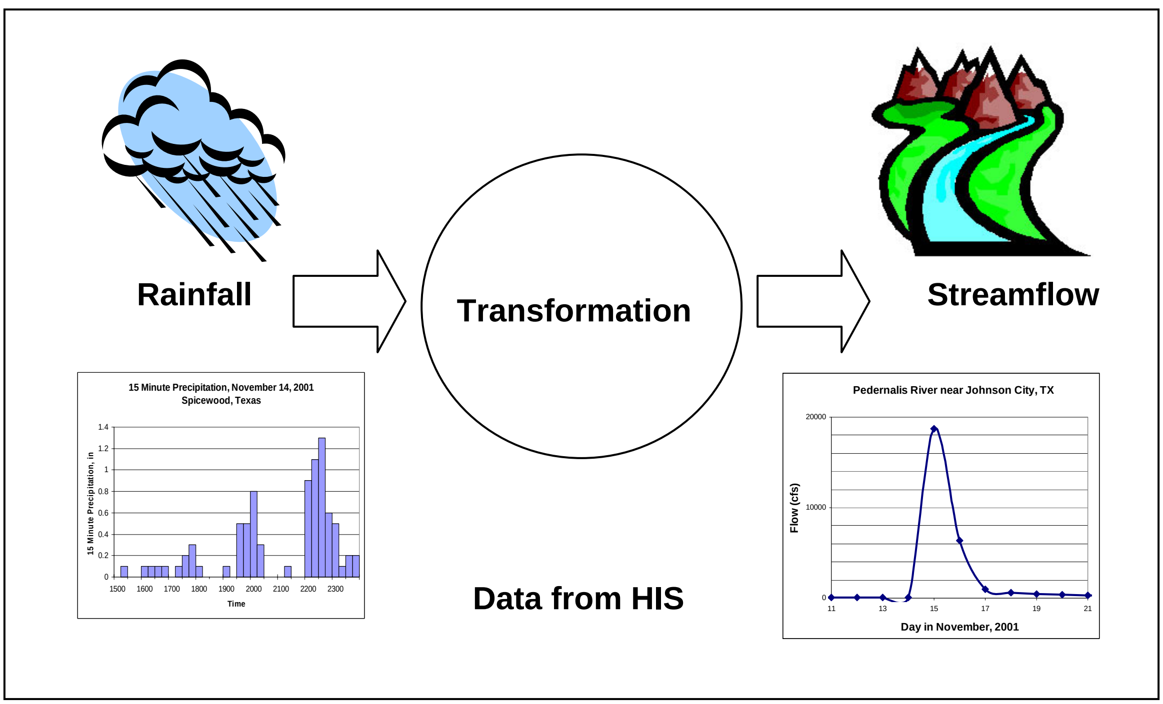


Figure 2.2 Precipitation to streamflow transformation

The ODM itself is a collection of a total of 27 tables and 147 fields (*Tarboton et al.*, 2007) comprising a relational database. This data model is not specific to any particular proprietary database software, creating a model that can be implemented at a range of scales, for a range of software applications. There are currently empty ODM databases and database schemas provided by CUAHSI in Microsoft SQL Server 2005 and Microsoft Access 2005 format. The collection of fields and tables that make up the ODM were designed to be comprehensive and flexible enough to hold all the essential information for a wide range of observation data sets, but also concise enough such that the database can be readily understood in its entirety.

The utility of such a data model is dependent on two factors: that it is a standard data model, and that it is used for many disparate data types and sources. While every attempt was made to make the ODM as comprehensive and flexible as possible within the constraints of a standard data model, some existing data models may not readily fit into the ODM. For such data sources there are two options. First, the ODM design can be modified to fit the individual data source. However, this option is undesirable because it departs from the standard data model concept. A dozen data sources stored in a dozen different customized versions of the ODM are only slightly more compatible than a dozen different data models. The second option is to leave behind specific pieces of information that do not have a matching field within the ODM. This is also undesirable as the loss of any piece of information may adversely affect future analysis. However, the ODM has been designed so that a permanent traceable heritage is held within the metadata. The information for any record in the ODM has a link to the original source. Thus, if certain information is deemed unessential at the time of transfer between the original source and ODM, that information can be retrieved through this traceable heritage. The specific fields in the ODM “VariablesCode” and “SitesCode”, both provide this traceable heritage. Neither of these two options is ideal. It is the responsibility of the entity transferring data to the ODM to choose the correct method of managing slightly incompatible data models.

The most recent information about the ODM, along with downloadable schemas and databases, can be found at http://www.cuahsi.org/his/odm.html.

2.2. HIS SERVER

The HIS Server is another component developed by CUAHSI's HIS program. HIS Server is a customized instance of ESRI's GIS Server, meant to act as both the front-end HIS map interface (see Figure 2.3), as well as the server that manages data download requests (see Figure 2.4). At the time of this document's creation, the HIS Server specifications were still in draft form. These draft specifications can be found in Appendix A. The most current specifications for HIS Server can be found at http://www.cuahsi.org/his.html. HIS Server is designed to be a standard that when used by any level of an HIS allows the user to browse observation points within a hydrologic geospatial context, and to use standard WaterOneFlow web services (see Section 2.3). Figure 2.4 shows an example of browsing National Water Information System (NWIS-purple circles) and EPA STORET (green triangles) southwest of Austin, Texas within the spatial context of roads and streams using HIS Server. Additionally, HIS Server has built in capabilities to allow the user to download data as a Microsoft Access Database or as a CSV file and to view selected data in a graph. As HIS Server continues to develop so too will the ability for individuals to access, browse, manipulate, and download hydrologic information.

One of the strengths of having the national HIS and multiple statewide HIS's use the same HIS Server specification is ability to share data. With only a few modifications, such as pointing a server to an individual data source, any HIS Server can access any HIS database.

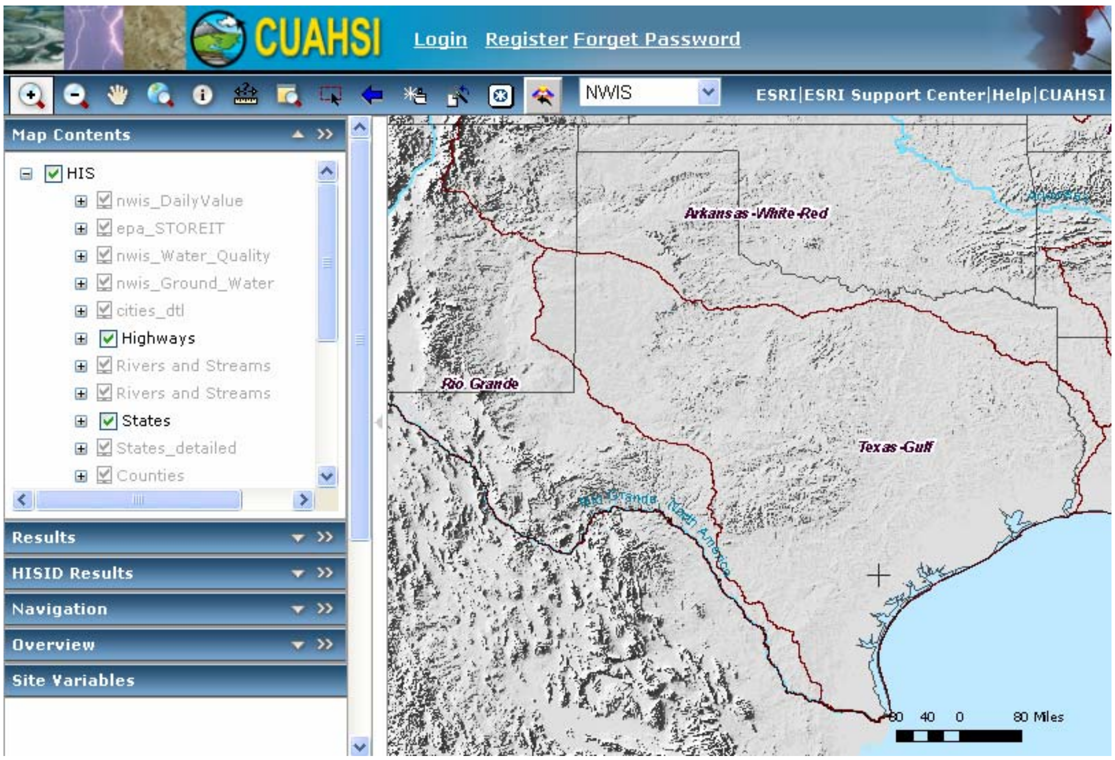


Figure 2.3 HIS Server prototype map interface

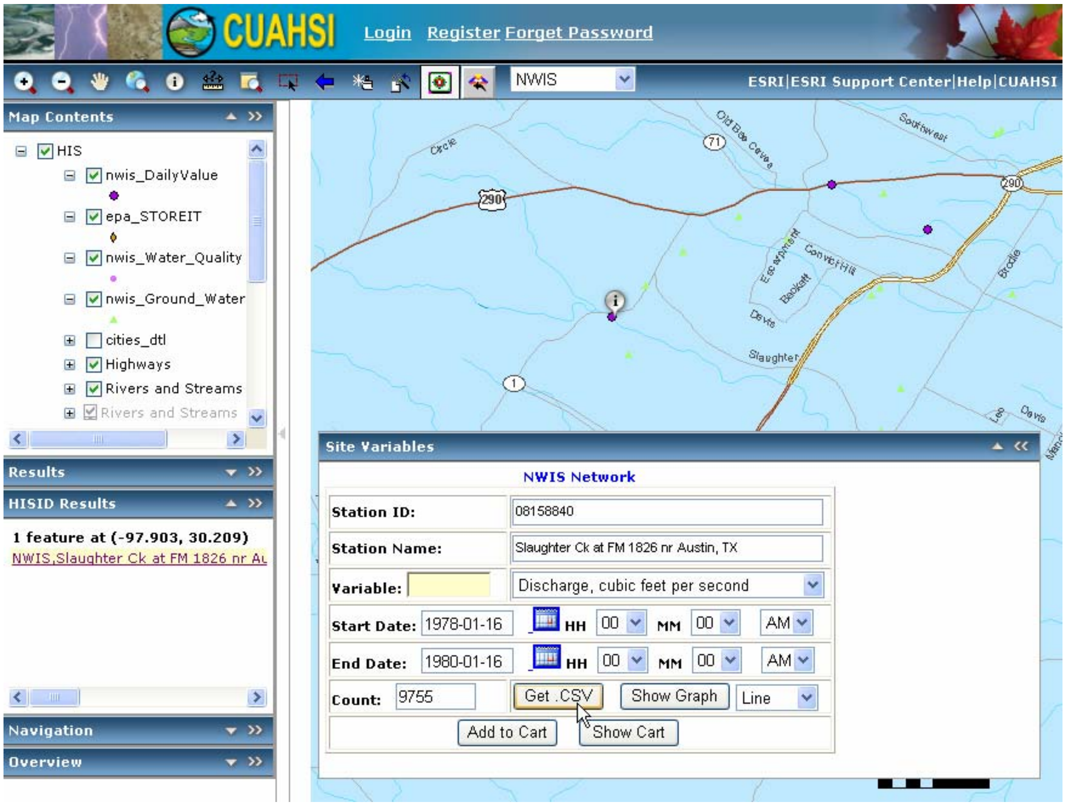


Figure 2.4 HIS Server map interface, close-up

2.3. WEB SERVICES

One of the key components of an HIS is the development of Web services. Web services have been defined as (*Booth et al.*, 2004):

> a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards.

Web services allow client computers to access remote data sources through a standard protocol, and return data requests in a standard format. This standard protocol is the Simple Object Access Protocol (SOAP). The advantage of a web service over other methods of data retrieval is that the standard protocol (the web service) can be hosted on a server as opposed to being hosted one each individual client. A more complete description of web services can be found in *Alameda*, 2006.

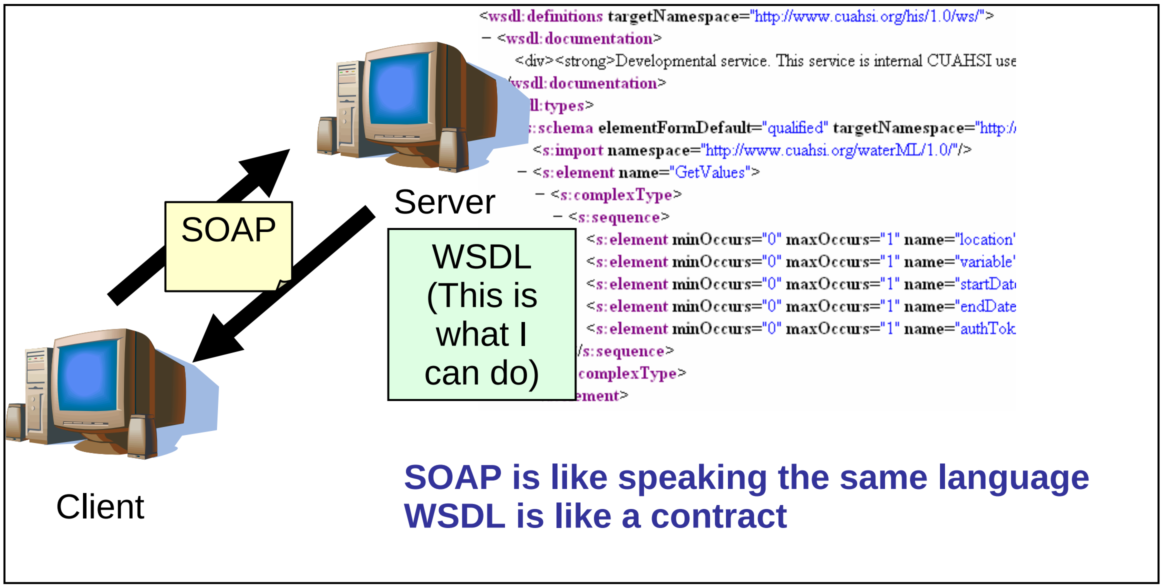


Figure 2.5 Web Services Diagram

The CUAHSI HIS development team has written a series of standard web services, called WaterOneFlow. The specific Web Service Definition Language (WSDL) for each of these services is defined at http://water.sdsc.edu/WaterOneFlow/, and includes at least three standard services for each of the data sources supported by WaterOneFlow. These three standard services are GetSiteInfo, GetVariableInfo and GetValues. These services are operational for the following data sources: NWIS, Daymet, MODIS, EPA STORET, and NAM 12k (see section 1.3 for descriptions of these and other data sources). Instructions for using these services have been compiled in the "CUAHSI HIS Web Services Workbook" (*Whiteaker et al.*, 2006).

Since the publication of the Web Services Workbook, a web service has been written for accessing data from any ODM database. Instructions for installing and using this web service, prepared by CUAHSI, can be found in Appendix B. With only a small amount of manipulation, such as directing the web services towards the proper server and ODM database, this web service can be easily customized to retrieve point observations data from any ODM data source.

These standard web services, when coupled with HIS Server, provide a framework for searching, accessing, and downloading hydrologic data. Slight modifications can be made to customize an HIS to statewide data sources. These modifications include creating new web services, or using existing ones (such as the ODM web service) to access new data sources. Further description of this customization is provided in Chapter 3 and Chapter 4.

2.4. NHDPLUS

The NHDPlus dataset is based on the National Hydrography Dataset (NHD) Medium Resolution (1:100,000 scale), which was originally developed by the United States Geologic Survey (USGS). More information about the NHD can be found at http://nhd.usgs.gov/. NHDPlus is an improved version of the 1:100,000 NHD that integrates the National Elevation Dataset (NED) and National Land Cover Dataset (NLCD). NHDPlus has been a cooperative project between the USGS and the EPA. The production of NHDPlus has been completed through a contract with Horizon Systems. More information about NHDPlus can be found at http://www.horizon-systems.com/nhdplus/. These data will eventually be provided directly by the EPA. A diagram of the NHDPlus file structure, as a geodatabase within ESRI's ArcCatalog is provided in Figure 2.6.

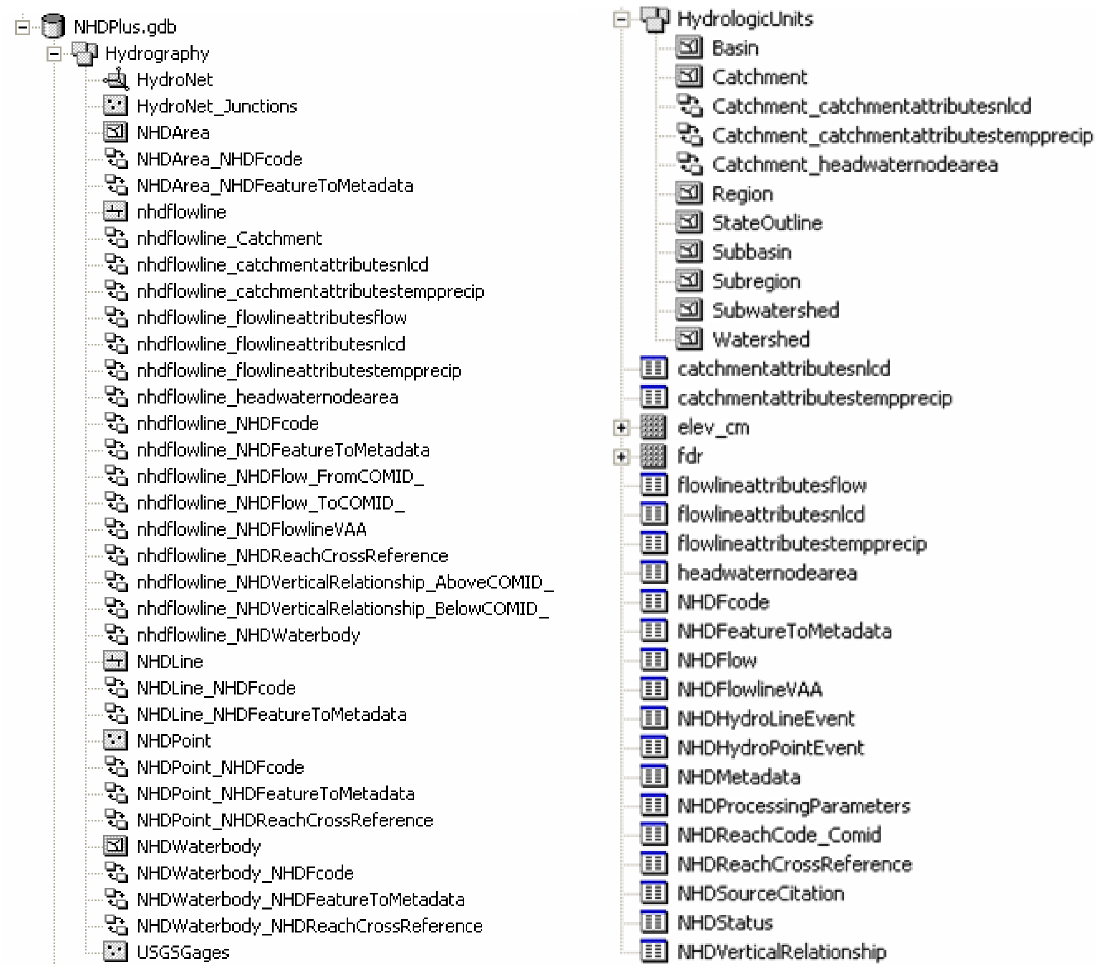


Figure 2.6 NHDPlus File Structure

A graphical example of NHDPlus can be seen in Figure 2.7. Beyond what is shown in this map, NHDPlus includes flow direction and flow accumulation grids, and numerous relationships between the features creating a broad picture of the hydrography of a region.

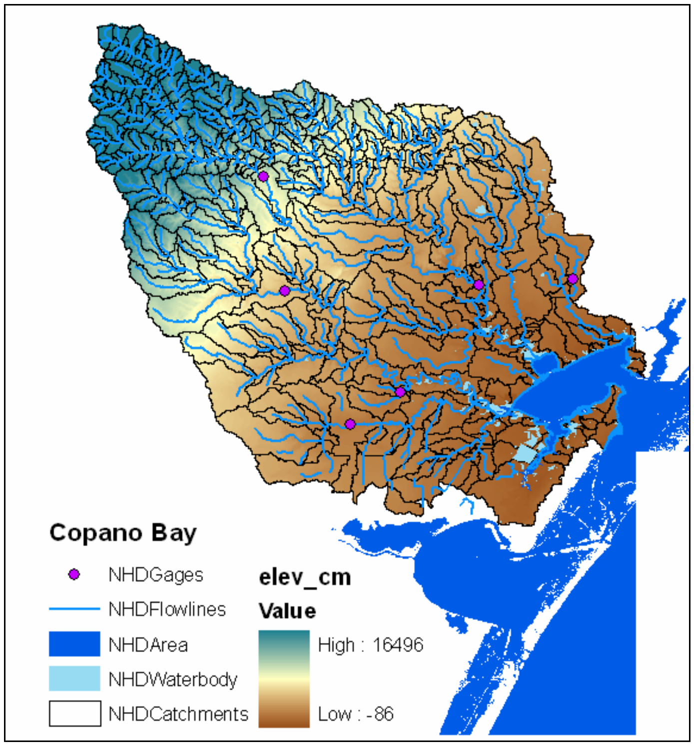


Figure 2.7 NHDPlus example

A large portion of the additional information found in NHDPlus is essentially the result of ArcHydro Terrain Preprocessing and Watershed Processing having been applied to the entire country. Elevation-based flow direction grids, flow accumulation grids, and catchments have been developed. The NLCD has been applied to these catchments, and information regarding catchment and watershed land cover linked to flowlines. Mean precipitation has been calculated for each catchment, and basic streamflow modeling has been conducted for each reach. Two methods have been used to determine mean annual flow and mean annual velocity. Thus, it is possible, without any additional calculations, to estimate stream flow and velocity at any given reach. Value Added Attributes, which assist with tracing throughout the stream network, have also been added. Lastly, USGS streamflow gages have been snapped to the network, allowing nearly seamless integration between NHDPlus and the NWIS (http://waterdata.usgs.gov/nwis). The snapped USGS Streamflow Gages are available through the USGS at http://water.usgs.gov/GIS/dsdl/USGS_Streamgages-NHD_Locations_GEODB.zip.

The NHDPlus is perfectly suited as a framework spatial reference for an HIS. Linking such a dataset to an HIS allows the time series hydrologic information to be discovered in the spatial context of rivers, streams, basins and catchments. Additional context such as relative stream size, land use and mean annual precipitation are also provided. Because an HIS is a collection of hydrologic data, it is logical to provide reference between these data and hydrologic spatial boundaries.

2.4.1. Literature Review

An HIS is major development in the field of hydroinformatics. This field extends far beyond the boundaries of HIS development. A recent comprehensive guide to this field is found in *Kumar et al.*, 2006. This text provides a background for the fusion of hydrology and information science. It is the product of numerous contributors, each specializing in a different aspect of hydroinformatics. Of particular interest to the context of the development of an HIS are the chapters on Hydrologic Metadata, Hydrologic Data Models, Data Formats, Web Services, Integrated Data Management System and Understanding Data Sources.

One important aspect relating to development of a statewide HIS is the specific data model recommended in this document: the ODM. This data model is not reviewed by Kumar et. al. Because the development of a statewide HIS is being coordinated in parallel with the development of a national HIS through CUAHSI, the data model used (and developed) by CUAHSI is also being used here. This model is most fully described in *Tarboton et al.*, 2006. The technical specifications of this model are reviewed above in Section 2.1.

The concept of a statewide HIS is not entirely new or unique. A similar effort has been produced and documented in *Soh et al.*, 2006. The Intelligent Joint Evolution of Data Information WebCenter for Hydroinformatics (http://water.unl.edu/) is essentially an HIS for the state of Nebraska. The WebCenter for Hydroinformatics provides a map-based interface to visualize data from a selection of groundwater and surface water sources. The sources are limited, and the functionality is limited to displaying graphs. Two key differences between the WebCenter for Hydroinformatics and the statewide HIS proposed in this document are the ability to actually download data (which is critical to integrating with hydrologic models) and the interaction with a national HIS, creating an HIS network.

Chapter 3. Methodology

One objective of this thesis is to provide documentation and instruction for the creation of a statewide HIS designed to operate in parallel with other information systems, such as a national HIS. While this documentation and instruction is meant to be used for any generic HIS, the specific example of a Texas HIS will typically be used to demonstrate these ideas. The research related to the writing of this document created a prototype for such an HIS for the state of Texas.

3.1. Preliminary Design

The Texas HIS is a network comprising of a single map-based web portal, connected to multiple data access servers. The specifications for the HIS Server have been developed by CUAHSI as part of the National HIS program, and are included in Appendix A. Interoperability between both the National, Local and Statewide HIS becomes possible if this server specification is used as the standard for statewide HISs and local observatories. Each HIS Server is connected to multiple individual data sources through a series of web services. A proposed design for use of the Texas HIS is given in Figure 3.1.

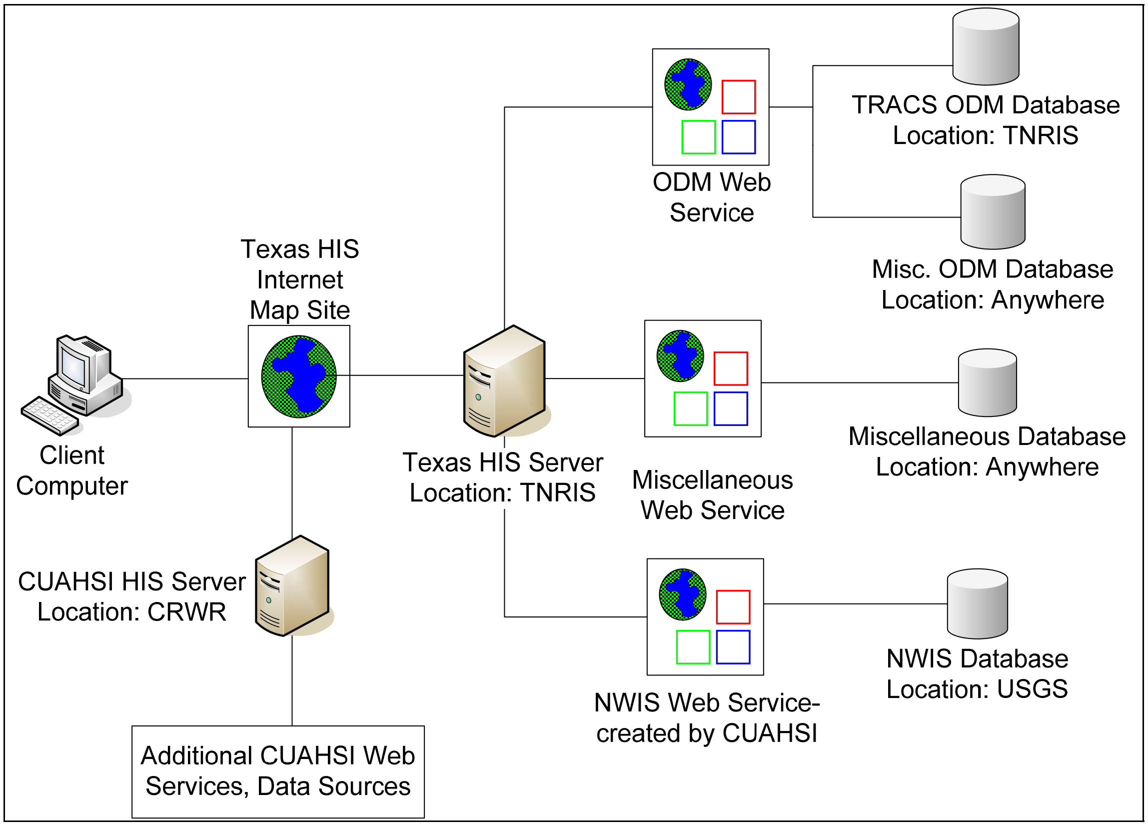


Figure 3.1 Texas HIS Network Diagram

3.2. DATABASE CONNECTIONS

As described above, the HIS Server is connected to a data source using web services. Among the web services used to connect the Texas HIS to its many data sources there are two primary methods: use of the ODM Web Service, and use of other miscellaneous web services. As part of the Texas HIS Prototype created as a component of this Thesis, the SWQM database has been connected to the Texas HIS. Additionally, the WIID has been considered for inclusion, but has not yet become part of the prototype. It is recommended that the WIID be one of the first additional sources added to the Texas HIS. See Table 1.2 through Table 1.5 for a list of these and other potential data sources and their descriptions.

The first method is to import data into an ODM database, for which web services are already built (see Figure 3.2). SWQM was connected using this method in the Texas HIS example. Importing data to the ODM can be performed using a SQL Server Integration Service (SSIS) package, or using the OD Data Loader, and is discussed in Section 3.4.

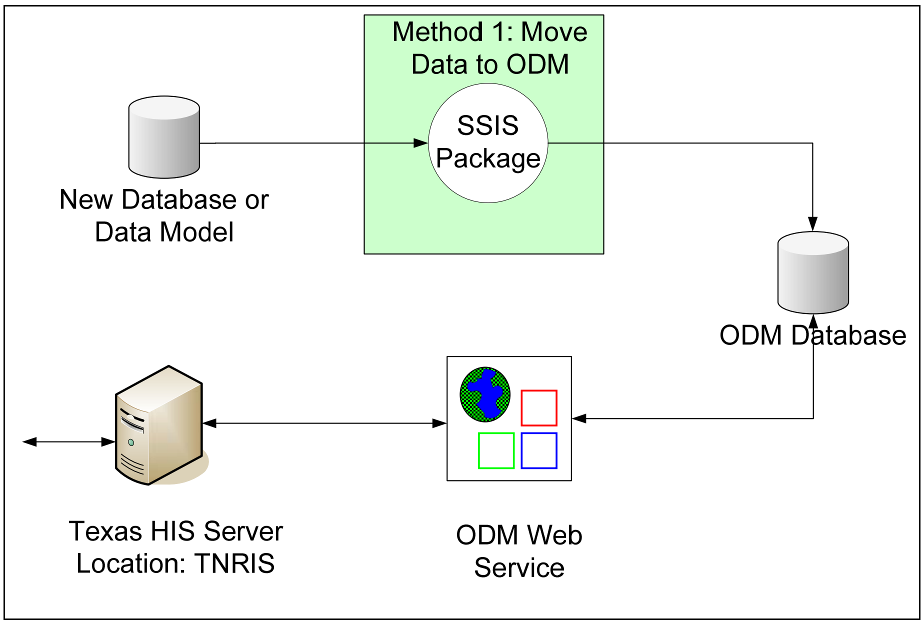


Figure 3.2 Method 1: Exporting Data to the ODM

There are two reasons for moving data from SWQM to ODM: 1) the SWQM database is not easily accessible by public queries, making it difficult to write a direct web service to the SWQM database from HIS Server, and 2) a generic web service for data stored in the ODM format has already been written as part of the CUAHSI National HIS.

For some data sources that already exist within a relational database, it may be possible to write a new web service to the existing database instead of migrating the data to a new database with an existing web service (see Figure 3.3). This would potentially be possible with the WIID. While schematically simpler and potentially more powerful, this method may be more complex because it requires writing a new web service, and should be reserved for connecting large-scale, established databases such as the USGS NWIS.

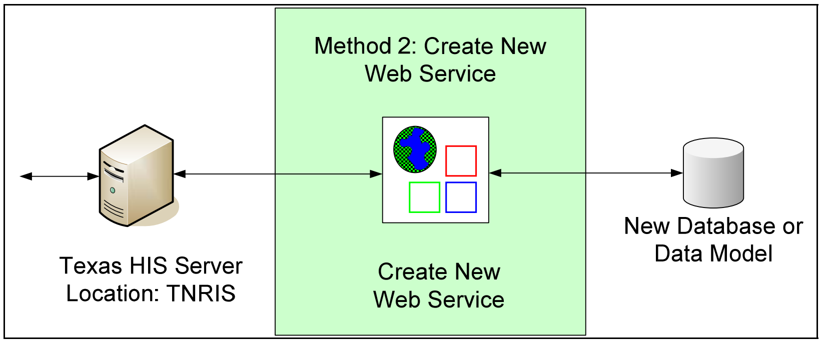


Figure 3.3 Method 2: Create New Web Service

3.3. Choice of the ODM as Preferred Data Model

While numerous data models exist, the ODM was chosen as the preferred data model for use with an HIS because it was created with that specific purpose in mind. It is comprehensive and flexible enough to store multiple different types of point time series information, and is becoming a standard among data models. Numerous applications and tools are being built upon the ODM through the CUAHSI National HIS program. Because the Texas HIS is utilizing the same data model as that being utilized at the national level, these same tools and applications can be used with data in the Texas HIS. It is highly recommended that future statewide HIS's incorporate data into the ODM in order to store and maintain data in a common format, advancing the goals of the HIS vision.

3.4. Loading Data into the Observations Data Model

Loading data into the ODM is an integral part of creating an HIS. Because web services and other common queries have been written for data in this schema, an HIS administrator does not have to create individual services for each new data source. Migration of data into the ODM is an example of an Extract, Transform and Load (ETL) process. Data is first extracted from its original form, transformed into the ODM form, and loaded into an ODM database.

There are two primary methods that can be used to load data into the ODM: 1) use the ready-built OD Data Loader tool created by CUAHSI and 2) create a custom built ETL process for an individual data source and database platform.

The OD Data Loader (also called ODM Data Loader) was created by the CUAHSI HIS development team, and is available at http://water.sdsc.edu/ODDataloader/ (see Figure 3.4). The OD Data Loader takes data from an Excel spreadsheet, tab delimited, or comma separated values file, and loads it into an instance of the ODM database in SQL Server. The limitation to this method is that the original file must have a specific configuration of columns to match with the loader. This configuration is a reduced version of the ODM. Data that does not fit into these fields cannot be loaded using OD Data Loader.

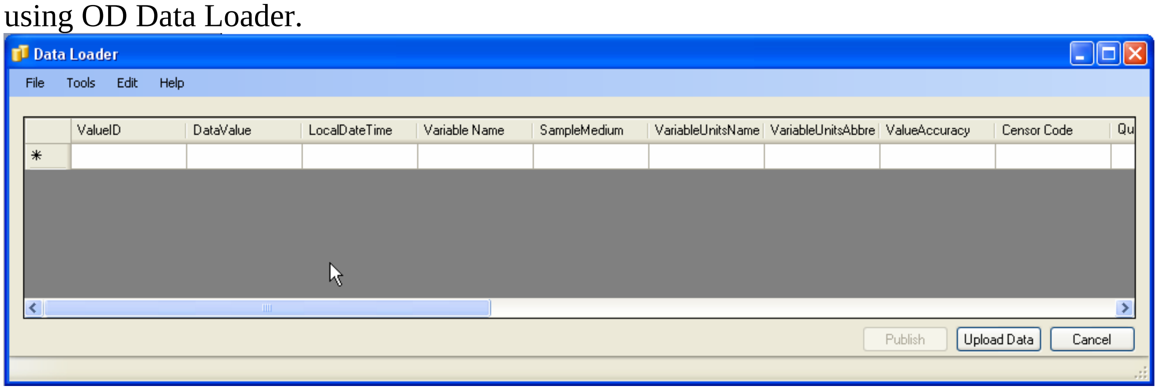


Figure 3.4 OD Data Loader

The specific fields that are included with the OD Data Loader are the following: ValueID, DataValue, LocalDateTime, Variable Name, Sample Medium, VariableUnitsName, VariableUnitsAbbreviation, ValueAccuracy, Censor Code, Quality Control, UTCOffset, QualifierID, QualifierCode, OffsetValue, OffsetTypeID, OffsetUnitsName, Site Code, Site Name, Latitude, Longitude, SRS Name and SampleID. While this is a long list of variables, it does not fill the entire ODM. Thus, some metadata that could be inserted into the ODM cannot be migrated using this tool.

Creating a custom ETL package takes more time and energy than using the OD Data Loader tool, but has the potential to enable a much more powerful data migration process. Use of such a package is not limited to the rigid data structure forced by the OD Data Loader. The architecture of the ETL package depends on the source data structure and the database platform used. To make SWQM accessible to the Texas HIS, data from the SWQM database was imported to an ODM Database using an ETL package created specifically for the SWQM data structure and for and ODM database in SQL Server 2005. In SQL Server, the ETL package is called a SQL Server Integration Service (SSIS). The following steps were used for this data migration:

- Fields of the SWQM database were mapped to the corresponding fields in the ODM Database.
- A SQL Server Integration Services (SSIS) package was developed to transfer the contents of SWQM to ODM
- The SSIS package was executed.

The example application in Section 4.4.3 provides a more complete description of this process, of the specific fields mapped from SWQM to ODM, and of issues that arose from this process.

3.5. SPATIAL CONTEXT

One of the great benefits of an HIS as a data discovery and analysis tool is that time series data is shown in its spatial context. The Texas HIS includes numerous two-dimensional GIS data layers to provide this spatial context (see Figure 1.1). While the ODM is adept at providing time series information and its associated metadata, the HIS Server map interface, developed on an ESRI® ArcIMS platform, is adept at providing the spatial context.

To create a spatial representation of the data points held in the ODM (the top three layers in Figure 1.1), essential fields from the ODM Sites table are used to create a series of points within a GIS. These fields can be obtained using a copy of the Sites table, or by using the GetSites call built into the ODM web services. Using the Latitude and Longitude fields of the Sites table as inputs, the "Make XY Event Layer" tool in ArcGIS creates this series of points. It is essential to include the SiteID field with the Event Layer in order to connect the Event Layer to the ODM. Thus, any time an Event Layer point is queried in the HIS Server map interface, information can be sent back and forth from the GIS to ODM using SiteID as the unique identifier for site information.

Adding additional two-dimensional spatial data as context for the ODM points is as simple as adding layers to an ArcMap document. Suggested layers include those that describe political boundaries such as National, State, County and City boundaries and those that describe natural features such rivers, lakes, mountains, geologic features and elevation (see Figure 1.1). It is useful to know that a water quality monitoring point is located along River A, just downstream of City B, upstream of Lake C, and flowing in the recharge zone of Aquifer E.

Not only is this spatial data useful as a context for discovering and investigating data sources, it is also useful for some modeling and analysis functions. Elevation and land use data is extremely useful for hydrologic modeling. The HIS Server development does not currently allow for the export of spatial data layers. This is a feature that is currently in development and should be available with future editions of HIS Server. This and other suggestions for further developments for HIS Server are included in section 4.6.

3.6. Adding Point Observation Layers to HIS Server

A key task in creating and maintaining an HIS is the addition of new datasets to the functionality of HIS Server. While the design of HIS Server is still being adjusted, a protocol for adding sources to this system has been developed by the CUAHSI HIS development team. The following steps describe this process:

1. Load data into a blank ODM instance. This can be done using ODM DataLoader, or by creating a customized SSIS package to migrate the data (described briefly in Section 3.2 and in greater detail in Section 4.4.3)
2. Copy the Web Services template to a new folder, edit the template web.config file to point to the new ODM, and test the web service to make sure it works as expected.
3. Create a point layer (a feature class or shapefile) from the new ODM "Sites" table, or from "GetSites" web service using the GetSites tool (described in Section 3.5)
4. Stop the HIS service
5. Add the point layer to the HIS .mxd map document, specify symbology, scale-dependent rendering, etc.
6. Add information about the new ODM, the associated web service, and the associated point layer, to HIS configuration file
7. Restart the HIS service

Figure 3.5 has been developed by the CUAHSI HIS development team to further describe this process in the context of a national HIS, with the numbers referring to the list above.

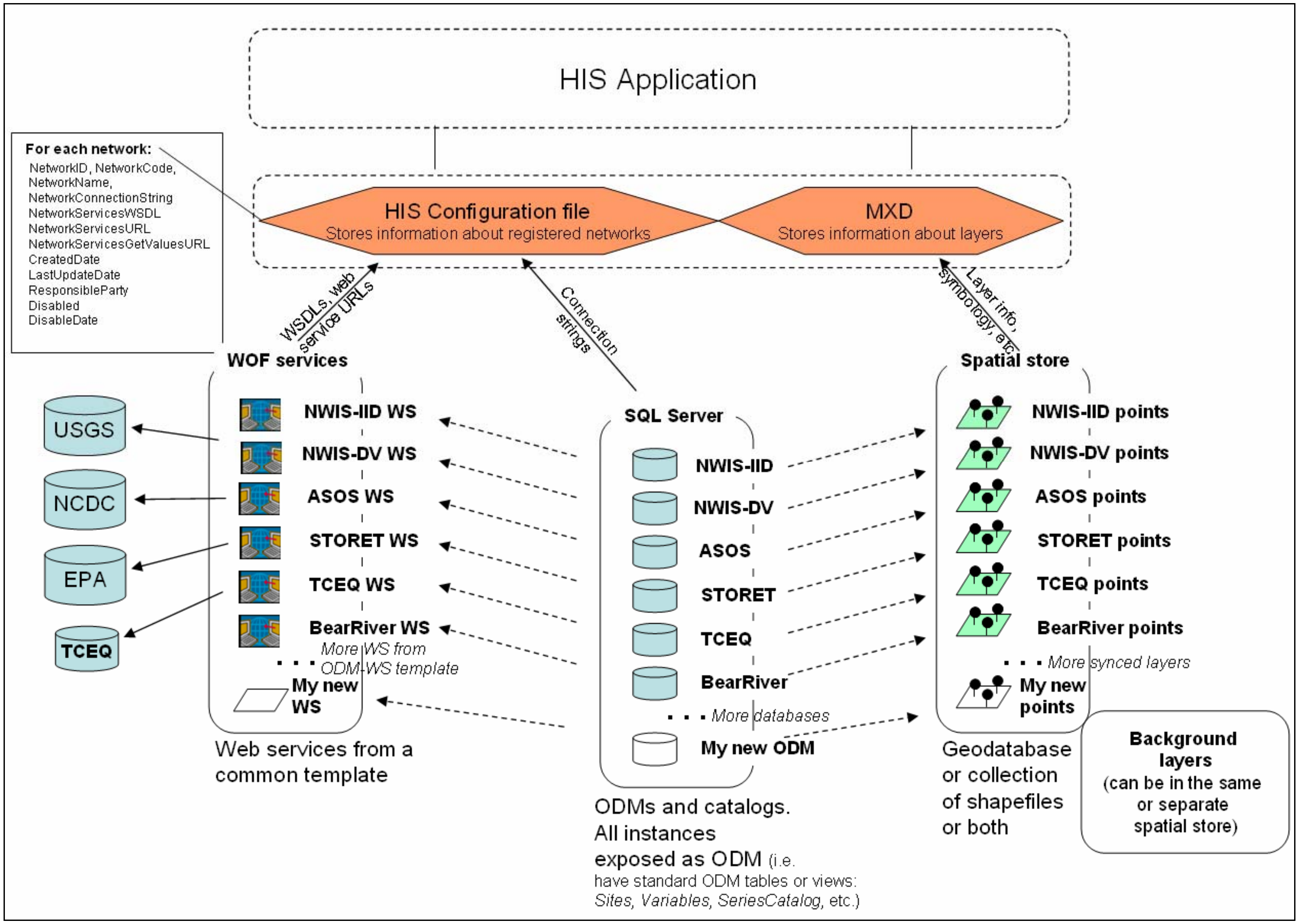


Figure 3.5 Adding New Data to HIS Server

3.7. DATA DISCOVERY

A key function of the HIS Map Interface is data discovery. Data discovery within this HIS design is limited to the functions built in to HIS Server. Data discovery within the ODM using direct SQL queries are possible, but require direct connection to ODM database. The following are methods that either currently exist as functions within HIS Server, or are suggested as functions to be built in to future editions. Other methods also exist.

3.7.1. Find sites by location

This method is used when a user already knows the location of the station from which data is desired. The current design of HIS Server requires the user to geographically navigate to the correct site, zooming in where necessary on the map interface (see Figure 2.3). Once zoomed in to an appropriate level, the site of interest becomes visible. This site can then be selected, and information regarding the variables collected and date range of data collection can be viewed. Data can then be added to the Data Cart, and downloaded as an Access Database (.mdb) or as a Comma Separated Values (CSV) file (see Figure 2.4). Tools to locate and zoom to a site automatically, given a site number or name are not currently available and should be considered for future development of HIS Server. The use of the SeriesCatalog as a metadata table allows ODM records to be easily queried to find the location of a site based on its site or name.

3.7.2. Find sites by variable

The ability to search sites by variable using the HIS Server interface does not currently exist. However, like searching by location, the use of the SeriesCatalog as a metadata table allows ODM records to be easily queried by variable. An example of such a use is selecting all sites that have data for the variable that corresponds to water temperature.

3.7.3. Find sample sites by number of records

The ability to search sites by number of records using the HIS Server interface does not currently exist. Again, the use of the SeriesCatalog as a metadata table allows ODM records to be easily queried by the number of records. An example of such a use is selecting only those sites that have more than 1000 water temperature records. Because many sites have only one or two observations that are often insufficient for many uses, it is important to be able to find only those sites with a large number of records.

3.8. DATA RETRIEVAL

Like data discovery, the data retrieval from the HIS is currently limited to functions built in to HIS Server. Additional methods of data retrieval are recommended for future editions of HIS Server and are currently being developed. Current capabilities allow the user to download data from selected sites for a specified period of time for a specified variable as a CSV file and also as a Microsoft Access database (.mdb). Additional methods of data retrieval concern the format of the Microsoft Access database that is downloaded from HIS Server. The current format is a geodatabase version of a simplified ODM. It is recommended that the option to download data in the full ODM format be developed.

3.9. INTEGRATION OF STATEWIDE HIS AND NATIONAL HIS

One of the strengths of using standardized data models and server structures like ODM and HIS Server for all levels of HIS is the ability to integrate information between these levels. This is especially true in the context of a national and a statewide HIS. All that is needed to add a national data source to a statewide HIS is to add the national source monitoring points to the statewide HIS Server, and copy the web service created by CUAHSI to the local HIS server (see Figure 3.6). When changes to the national data source access protocol occur, CUAHSI updates the specific web service and releases it to each of the HIS Server instances. It should be noted that this system requires significant communication between HIS instances and the CUAHSI HIS development tem. Similarly, a statewide data source can easily be incorporated into the national HIS by adding the statewide data source observations points and a copy of the local data source web service. In this way situations where national data has holes in it, such as EPA STORET and TCEQ SWQM, can be fixed (see Figure 1.8, Figure 1.9 and Figure 1.10).

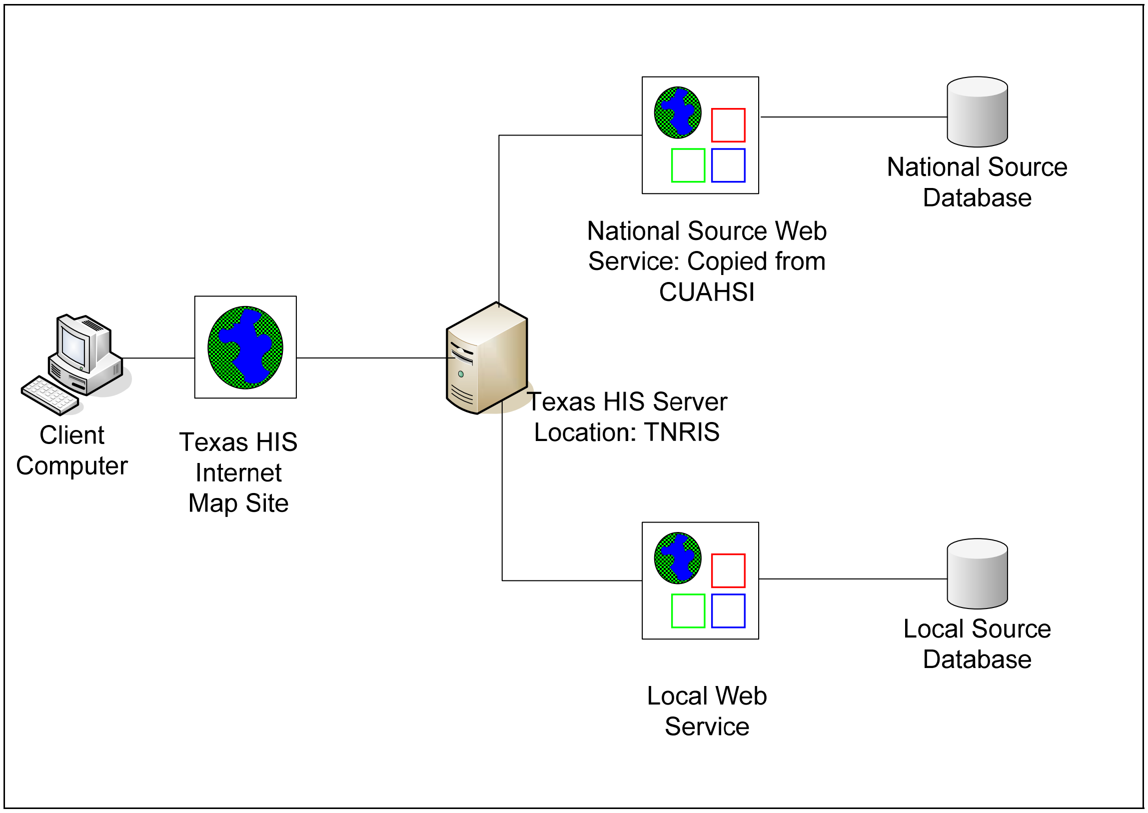


Figure 3.6 Integration of Statewide HIS with National data source

3.10. TECHNICAL LIMITATIONS

The development of HIS is a work in progress. As such, numerous technical limitations currently exist for which solutions are currently being sought. It is recommended that before implementing a state HIS, the most current specifications for HIS Server and ODM are referenced through the CUAHSI HIS program webpage http://www.cuahsi.org/his.html. This dynamic state of HIS Server, and to a lesser extent ODM, can be seen as a technical limitation for entities implementing this technology. Any deployment of HIS Server will most certainly be faced with upgrades and new versions. However, the nature of cutting edge technology such as the HIS is one of improvement and iteration.

3.11. INSTITUTIONAL LIMITATIONS

The development of a statewide HIS is dependent on cooperation and data sharing amongst numerous groups and agencies. Such cooperation requires communication and coordination on the part of the organizing agency.

The level of cooperation and communication required depends on the current availability of the hydrologic information. In some cases (such as that of SWQM), the data is already available to unrestricted users over the internet. This allows the coordinating agency to download, scrape, or develop web services to this available data. Although communication and coordination between the source agency and the coordinating agency is not required, it is highly recommended. Downloading or scraping data can create a large burden on the host server. Coordinating the scraping with the source agency may relieve some of this burden, and speed up file transfer. Also, automatic scraping and web service codes are easily broken by a small change in the source data access method. Through proper coordination, such changes can be communicated to the organizing agency prior to such a tool breaking. If the nature of such changes is properly communicated, significant inconvenience to the user can be avoided.

Data that is not readily available over the internet can be much more difficult to add to an HIS. Addition of such data always requires inter-agency cooperation and communication. This can be as simple as a data request and the transfer of a few files on a CD, or as complicated as directly connecting the source database server to the HIS Server through a custom-built web service. Regardless of the method chosen, communication is required between the organizing and the source agencies. If the specific source data is periodically updated, than protocols and agreements for the transfer of new data must also be made.

Depending on the existing relationship between the organizing and the source agency, such communication and coordination may be quite difficult. Internal agency politics often come into play. Many data owners are hesitant to share their data for fear that it could be misused or misrepresented. In these respects the addition of each new data source must be considered individually and appropriate action taken. Such is the limitations imposed by the institutions, agencies, and groups who own the data.

3.12. RECOMMENDATIONS FOR FUTURE WORK

As with many research projects, the development of the HIS framework and technology, both for a national and a statewide edition, is still being completed. What has been described in this document is only an iteration of a product that has many potential improvements and additions. The following is a summary of some of the areas where these improvements to the HIS framework may take place. Many of these issues are currently being addressed by the HIS workgroup within CUAHSI.

3.12.1. Observations Data

After four major pre-release editions, the ODM has been released as Release Version 1.0. See http://www.cuahsi.org/his/odm.html for the most recent update information. The iterations leading up to this final product have led to numerous essential changes, improving the flexibility and uniformity of this data model. With an official release edition, users are beginning to migrate large volumes of data into the ODM 1.0 format. As more data becomes available in this model, remaining model weaknesses may be discovered. As additional editions are released, tools that convert old databases to the new format need to also be created. Additionally, any tools that use ODM data as an input may need to be adjusted. One example of such a tool is the web service that connects HIS Server to ODM data. Such changes have the potential to be time and energy intensive. Thus, the advantages of potential changes to the ODM need to be weighed against the work necessary to respond to those changes. Additionally, communication with the ODM user community is paramount.

For the utility of the ODM to increase, various users need to move existing datasets to this format. While bureaucratic inertia may impede such progress, it is recommended that agreements between organizations and agencies that collect and store data bed made. Additionally, new data should be collected and stored in this format. The impetus for this data migration is a bit circular. As more data is moved to the ODM, the wide use of the ODM increases. However, many data managers will only take the time to move data if the format is already widely used. One reason for moving the SWQM data to the ODM as part of the Texas HIS is to begin this data migration process and increase the use of ODM.

Another factor that increases the utility of the ODM is the creation of tools for ODM data manipulation. Such tools might include capabilities to graphically display the sampling density of a particular variable for a particular date range, identify stations with the longest date range, and even symbolize the map by average value of a particular variable. It is recommended that future research concentrate on this task. Publishing commonly used queries as well as making ODM data easier to access, browse, and download is an essential task.

3.12.2. HIS Server

Like the ODM, HIS Server is an emerging technology, with expected improvements and iterations still to come. Increased use will shed light on existing issues, such as scaling and stability. Also like the ODM, there is room for additional tools to be built on top of HIS Server.

One such tool that should be developed is one that allows the user to select and download geospatial data. Currently, HIS Server provides the capability to select an observation point, and download data for a given variable and period of record. However, a significant portion of hydrologic information is stored as two, three, and four dimensional geospatial data. The ability to select and download geospatial data (and its tabular attibtutes) based on a given shape would be extremely useful. This should begin with two-dimensional data. Additionally, tools that allow for the selection of an area based on watershed properties would also be useful. An example of this is being able to select an entire upstream watershed, given a point on a river network, and download the corresponding hydrologic information. A similar watershed selection tool is currently in production by the USGS at http://water.usgs.gov/osw/streamstats/ssonline.html. As three and four-dimensional data (such as NetCDF) is better understood, tools that allow the visualization and access to such data within the HIS Server framework should also be developed.

Additional tools that should be developed involve the manipulation of HIS Server data. While the current graphing utility works well, numerous improvements can be made. Adding statistical tools such as those that exist in the NWIS Analyst tool created at Utah State University (http://water.usu.edu/nwisanalyst/) would be a significant improvement. The NHDPlus features and navigation tools could also be mobilized for efficient network tracing and watershed characterization within the HIS Server framework.

Additionally, a series of tools to enable smarter selection of data (as discussed in section 3.7) should also be developed. Such tools should allow quick navigation to sites that measure a particular variable, and quick navigation by other site attributes such as site number, number of variables measured, and number of values measured.

Chapter 4. Texas HIS

A major goal of this project was to create a prototype of a statewide HIS in the state of Texas: a Texas HIS. The Texas HIS prototype was created in order to demonstrate many of the methodologies discussed in Chapter 3. The Texas Natural Resources Information System (TNRIS), a division of the Texas Water Development Board (TWDB) and responsible among other things for maintaining hydrologic information, is the logical host for such a data portal. See http://www.tnris.state.tx.us/ for more information on TNRIS. The server infrastructure at TNRIS already exists, as well as many of the statewide data sources to be included in the Texas HIS. Creation of a Texas HIS also provides for the demonstration of communication between a national HIS (through CUAHSI) and a state HIS (through TNRIS). By partnering, data can be shared between the two systems and a more complete hydrologic picture created. Many of the examples used to demonstrate the methods discussed in Chapter 3 are in fact examples from the creation of a Texas HIS prototype. The following section discusses the creation of this prototype system in more detail.

4.1. PROTOTYPE

With the plethora of available hydrologic data, the extent of a statewide HIS is only limited by time, effort, and computational space needed to host such a project. The Texas HIS prototype has been completed with a small amount of data, with the intent that additional data sources are steadily added. Section 3.2 discusses the connection to additional point data sources. The Texas HIS prototype presented here includes one data source within the context of multiple spatial data layers. The spatial layers to be included are discussed in Sections 4.2 and 4.3. Additionally, a discussion of the surface water quality point observations to be included is discussed in Section 4.4.

4.2. HIGH RESOLUTION NHD

One of the key spatial features for the Texas HIS is the High Resolution NHD (NHDH or NHD24k). High Resolution NHD is mapped at the 1:24,000 scale, as opposed to Medium Resolution NHD which is mapped at the 1:100,000 scale. Both datasets can be accessed at http://nhd.usgs.gov/index.html. Because of the higher spatial resolution, the NHDH has a much higher stream density than the Medium Resolution NHD, and is thus prone to network errors and disconnected streamlines that prevent accurate flow tracing. Before inclusion with the Texas HIS, the connectivity of the NHDH for Texas was analyzed and major errors manually corrected. The following section describes the error correction process. As part of the creation of the Texas HIS prototype, the edited version of NHDH was delivered to TNRIS as an SDE database feature class.

The use of error-free datasets is essential to the utility and usability of the Texas HIS. As datasets are added to an HIS they should be checked for errors and consistency with other data.

4.2.1. Introduction

Part 2 of Phase One of the contract between the Center for Research in Water Resources (CRWR) and the TWDB calls for a quality assurance review and updates to the NHDH as part of the StratMap program. This report details the completion of that task, the methodology employed, and the specific changes made to the dataset.

The NHDH is a comprehensive set of digital spatial data with information about surface water features (*USGS*, 2007) originally produced by the USGS. It includes spatial information about rivers, streams, lakes and reservoirs, all of which create a network of surface water features in the United States. The NHDH was created at the 1:24,000 scale, and has a significantly denser stream network than the Medium Density NHD, produced at a 1:100,000 scale. One of the primary utilities provided by the NHDH is the ability to create a digital stream network, and to trace flow from the top to the bottom of a watershed. This utility is dependent on the complete connectivity of the streamlines in the network.

The NHDH is a compilation of many previous versions of surface water spatial information, including the USGS Digital Line Graph hydrography data and the Environmental Protection Agency Reach File 3. Further information regarding NHDH specifications, as well as the data itself, can be found at http://nhd.usgs.gov. Due to the high resolution nature of the NHDH, and the inability to manually verify the spatial accuracy of the more than 2 million individual reaches, there are numerous breaks in the connectivity of the NHDH.

For cataloging and organizational purposes, the NHDH is subdivided into smaller and smaller units, each with an additional series of digits to describe the area with greater precision. For instance, the entire United States has been divided into 21 2-digit hydrology units (HUC) called regions. Each of these 2-digit HUC's is divided into multiple 4-digit HUCs called subregions, which are then subdivided into 6 and 8 digit areas. The NHDH is made available by subregion. The State of Texas is comprised by portions of 24 subregions (See Figure 4.1).

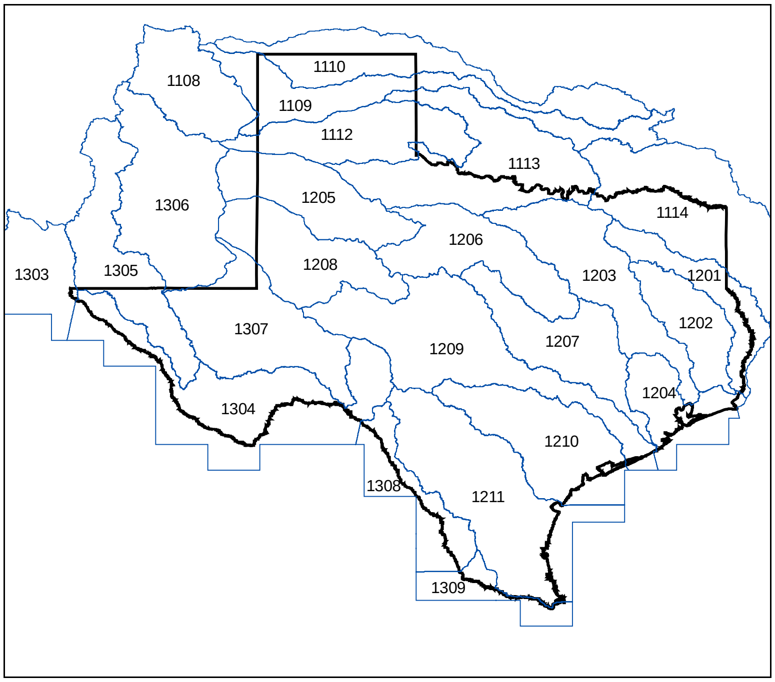


Figure 4.1 Subregions in Texas

Within these 24 subregions, thousands of digitized streams are not connected to the streams into which they should flow. These disconnects represent an error in the NHDH where for some reason the spatial representation or surface water does not accurately represent physical surface water flow. The high number of such disconnects makes it impossible to edit and fix each one. However, a few of these disconnected portions of the NHDH are positioned such that if fixed, they would dramatically increase the connectivity of the entire NHDH. For instance a small break halfway down a major river disconnects the entire upper watershed from the entire lower watershed. If fixed, the entire upper watershed is reconnected. These critical disconnects are the focus of the NHDH portion of the Texas HIS project.

4.2.2. Purpose

The purpose of this section is to determine the locations of critical breaks in the NHDH within Texas, and to fix these disconnects using existing geospatial tools as a guide. Fixing these critical disconnects dramatically increases the connectivity of the network, and increases the utility of this spatial product to multiple other users. This process involved first locating these disconnects and then fixing them in a way such that the updated data may be utilized by other users, and that the nature of the edits is recorded.

4.2.3. Methodology

The methodology used to complete this task can be divided into two distinct parts: locating critical disconnects, and correcting critical disconnects.

4.2.3.1. Locating Critical Disconnects

Because the NHDH for the entire state of Texas is too large to easily build a geometric network, the following analysis was conducted on a subregion by subregion basis. Each of the 24 subregions (see Appendix C.1) was downloaded individually from the NHD ftp download site (ftp://nhdftp.usgs.gov/SubRegions/) in a geodatabase format. The connectivity of the network of each of these subregions was then analyzed individually. The primary tool used to locate critical network disconnects was the ArcGIS Network Analyst Trace Task toolbar. By tracing both upstream and downstream, critical disconnects were located. To simulate “fixing” the critical disconnects, additional network tracing flags were placed (See Figure 4.2). Thus, the number of errors is the same as the number of additional network tracing flags placed.

i) ii)

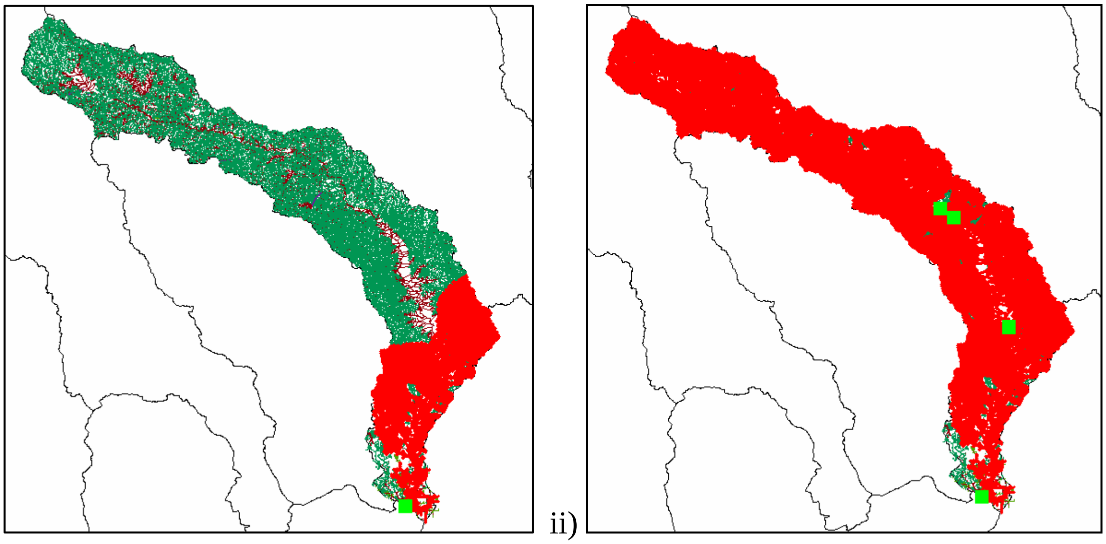
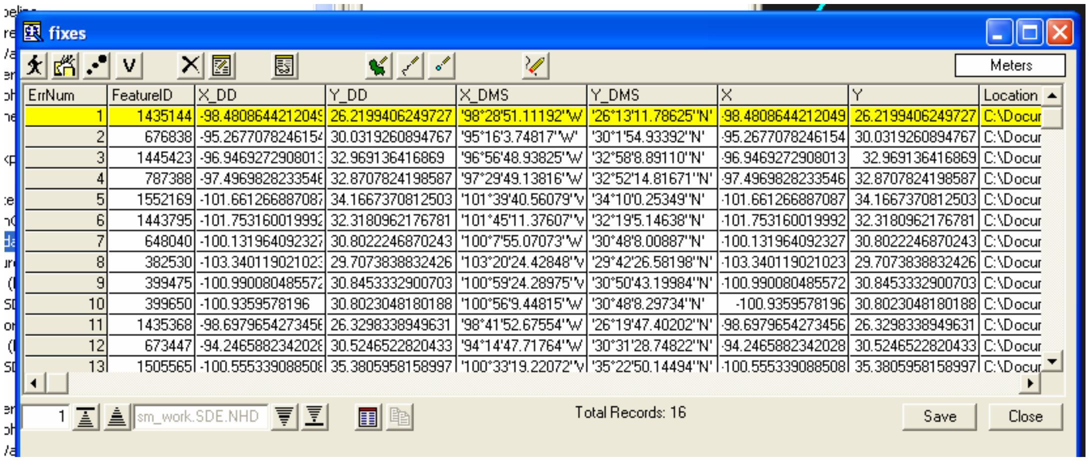


Figure 4.2 Discovering disconnects in the NHDH: i) A disconnect in Subregion 1201 prevents the entire subregion from being captured in a network trace (shown in red). ii) The addition of three flags, simulating three fixes, increases the network connectivity.

The location of each error was recorded for future editing using the COMID as the location identifier. The COMID field in the NHDFlowline feature class is the unique identifier for all flowlines, and is thus the logical reference for location. Locating the critical disconnects was performed on the computers at the CRWR using NHDH geodatabase data downloaded as Subregions from the NHD ftp site in September, 2005.

4.2.3.2.Fixing the Critical Disconnects

Once the critical disconnects were located in individual subregions, the same disconnects were relocated within the entire Texas NHDH on the TNRIS system. Using the Production Line Tool Set (PLTS) extension within ArcGIS, the location of each of the previously determined critical disconnects was reestablished, and bookmarked in a PLTS Data ReViewer table (see Figure 4.3). This table provides easy navigation from error to error in order to facilitate editing. See Appendix C.3 for a list of the COMIDs edited.



ErrNum	FeatureID	X_DD	Y_DD	X_DMS	Y_DMS	X	Y	Location
1	1435144	-98.480864421204	26.2199406249727	'98°28'51.11192"W	'26°13'11.78625"N'	-98.4808644212049	26.2199406249727	C:\Docur
2	676838	-95.267707824615	30.0319260894767	'95°16'3.74817"W'	'30°1'54.93392"N'	-95.2677078246154	30.0319260894767	C:\Docur
3	1445423	-96.946927290801	32.969136416869	'96°56'48.93825"W	'32°58'8.89110"N'	-96.9469272908013	32.969136416869	C:\Docur
4	787388	-97.496982823354	32.8707824198587	'97°29'49.13816"W	'32°52'14.81671"N'	-97.4969828233546	32.8707824198587	C:\Docur
5	1552169	-101.66126688708	34.1667370812503	'101°39'40.56079"V	'34°10'0.25349"N'	-101.661266887087	34.1667370812503	C:\Docur
6	1443795	-101.75316001999	32.3180962176781	'101°45'11.37607"V	'32°19'5.14638"N'	-101.753160019992	32.3180962176781	C:\Docur
7	648040	-100.13196409232	30.8022246870243	'100°7'55.07073"W	'30°48'8.00887"N'	-100.131964092327	30.8022246870243	C:\Docur
8	382530	-103.34011902102	29.7073838832426	'103°20'24.42848"V	'29°42'26.58198"N'	-103.340119021023	29.7073838832426	C:\Docur
9	399475	-100.99008048557	30.8453332900703	'100°59'24.28975"V	'30°50'43.19984"N'	-100.990080485572	30.8453332900703	C:\Docur
10	399650	-100.9359578196	30.8023048180188	'100°56'9.44815"W	'30°48'8.29734"N'	-100.9359578196	30.8023048180188	C:\Docur
11	1435368	-98.697965427345	26.3298338949631	'98°41'52.67554"W	'26°19'47.40202"N'	-98.6979654273456	26.3298338949631	C:\Docur
12	673447	-94.246588234202	30.5246522820433	'94°14'47.71764"W	'30°31'28.74822"N'	-94.2465882342028	30.5246522820433	C:\Docur
13	1505565	-100.55533908850	35.3805958158997	'100°33'19.22072"V	'35°22'50.14494"N'	-100.555339088508	35.3805958158997	C:\Docur

Figure 4.3 PLTS Data ReViewer Table

With all the errors bookmarked, each disconnect was compared with additional spatial data to determine the correct course of action for editing. Depending on the error and its location, these additional datasets included Digital Orthophoto Quarter Quad (DOQQ) aerial imagery, mosaiced USGS Digital Raster Graphic (DRG) maps, NHDH network in geodatabase format, 30 m NED and ESRI Geography Network Satellite Imagery. Appendix C.2 gives the specific location for each of these sources. An obvious difference between the aerial imagery and the NHDH flowline existed for most of the disconnects discovered using network tracing tools. The common case was a flowline that was not properly connected to a junction. See Figure 4.4 for an example of such an error.

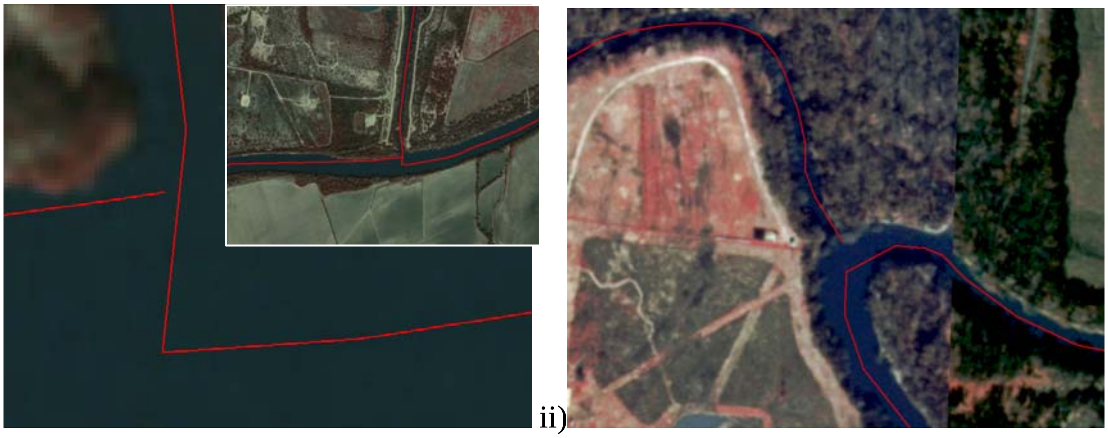
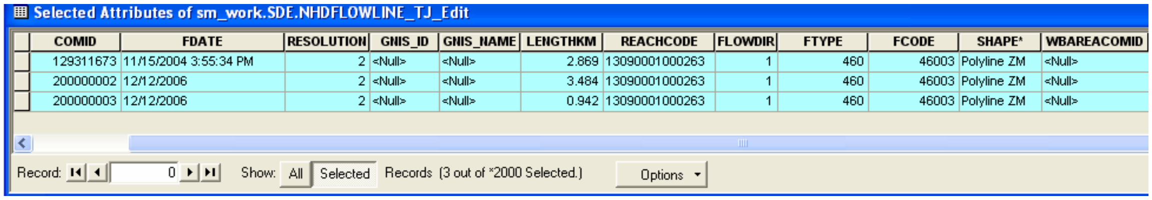
i)  ii)

Figure 4.4 Disconnect errors in the NHDH: Two examples of common disconnect errors. A simple adjustment in flowline geometry would fix the error. Also, an adjustment is verified by the DOQQ imagery in the background.

Another common case was an improperly set flow direction. This was fixed by changing the flow direction, and verifying the fix with both the NED and the DRGs. A few of the errors did not have clear fixes in the available aerial imagery. In these cases the error was reviewed by representatives from both TNRIS and CRWR, and a solution was agreed upon.

As NHDH features were edited, adjusted, deleted, and added, a record of changes was kept in the NHDFeatureToMetadata table. This table records which edits were made to which features, and is a link between the NHDFlowline feature class and the NHDMetadata table via the unique indentifier COMID (for the feature) and DUUID (for the metadata record). This editing process created new NHDFeatureToMetadata and NHDMetadata tables in the exact same format as those already used in the NHDH. See Appendix C.4 for these two tables.

Some editing required the creation of new flowlines. These were assigned a COMID beginning with 200000001. This COMID was chosen because it is higher than the highest COMID in the NHDH for Texas, and is thus unique. The remaining attributes for these new flowlines were copied from the neighboring flowline, and include Reachcode, FType, FCode, GNIS_Name and GNIS_ID (See Figure 4.5)



Selected Attributes of sm_work.SDE.NHDFLOWLINE_TJ_Edit

COMID	FDATE	RESOLUTION	GNIS_ID	GNIS_NAME	LENGTHKM	REACHCODE	FLOWDIR	FTYPE	FCODE	SHAPE*	WBAREACOMID
129311673	11/15/2004 3:55:34 PM	2	<Null>	<Null>	2.869	13090001000263	1	460	46003	Polyline ZM	<Null>
200000002	12/12/2006	2	<Null>	<Null>	3.484	13090001000263	1	460	46003	Polyline ZM	<Null>
200000003	12/12/2006	2	<Null>	<Null>	0.942	13090001000263	1	460	46003	Polyline ZM	<Null>

Record: 0 Show: All Selected Records (3 out of *2000 Selected.) Options

Figure 4.5 New feature attributes

After all the editing was complete, the LengthKM attribute was updated for all the changed and added flowlines based on the length when projected in the Albers Equal Area projection, based on the North American Datum of 1983. The specific projection information is: Standard Parallels = 29º30' N and 45º30' N, with a Central Meridian at 96º W and the Projection Origin at 23º N. This projection is the same as that used by the National Hydrography Dataset (*USGS*, 2000).

4.2.4. Results and Deliverables

The result of this task is an improved version of the NHD24k for Texas. The deliverables include:

- An edited NHDFlowline featureclass in SDE Format. The path for this file is: Huan.sm_work.SDE.NHDFLOWLINE_TJ_EDIT
- A geodatabase containing two essential metadata tables for the edited data:
  - NHDFeatureToMetadata
  - NHDMetadata
- A progress report, provided in both paper and digital formats, including appendices describing the extent of edits performed on the NHDH in Texas, and describing the methods and processes used.

4.3. NHDPlus

Discussed as a national data source in section 2.4, the NHDPlus is an integral portion of the Texas HIS. While somewhat redundant with the NHDH, the NHDPlus data provides additional information about the watershed, albeit at a lower resolution. This additional information includes modeled mean annual stream flow, stream velocity, land use statistics, and essential network navigation features. Depending on the specific HIS users' needs, NHDPlus or NHDH may be a more appropriate spatial context. While HIS Server features do not currently allow for a full utilization of NHDPlus' capabilities, it is hoped that future versions of the server tool will.

4.4. Surface Water Quality Monitoring data

The primary data source included in the Texas HIS prototype is the Surface Water Quality Monitoring (SWQM) data made available by the TCEQ through their TRACS database system. The SWQM data is often referred to by its parent database, TRACS, even though TRACS contains more than just SWQM data (*Trujillo*, 2006). In this document, an effort has been made to only use the term TRACS when referring to the entire TCEQ database, and the term SWQM when referring to the specific surface water portion of that data. While available in an online data access system, the SWQM data is not easy to query. Because of this, the SWQM data was migrated from the TRACS database to an ODM 4.0 database hosted by TNRIS. Various difficulties in transferring data from SWQM to ODM have been encountered due to differences between the overall data models. These difficulties have been discussed with the creators of the ODM, various solutions proposed, and a final solution adopted for the migration of this data, which was completed with the expert help of Chris Williams, the TNRIS Database Administrator. This section is an attempt to describe these difficulties, possible solutions, and the final solution that was adopted.

4.4.1. Background

The data that is incorporated as part of Texas HIS from SWQM is time series point observations describing water quality in Texas.

The TCEQ is currently in process of updating the SWQM dataset within the TRACS database to an easier to use web interactive database, called Surface Water Quality Monitoring Information System (SWQMIS). However, in an effort to complete a prototype of the Texas HIS by May, 2007, and to better understand the relationship between ODM 4.0 and SWQM, the old version of SWQM is being used for this current application.

Like the SWQM database, the ODM is in a state of improvement and validation. This section will look specifically at ODM 4.0, as described in *Tarboton et al.*, 2006. ODM 4.0 is a pre-release edition of the model. Since this data migration project was initiated, a full release version of the model, ODM Release Version 1.0, has been released. Despite the existence of this more recent version, the description of the migration of SWQM data to the ODM refers to ODM 4.0 in order to maintain consistency.

The goal of this data migration application was to move all the data from the old version of SWQM to an ODM 4.0 database that is served at TNRIS as part of the TxHIS. One method of loading the data into ODM 4.0 is the OD Data Loader (also called ODM Data Loader) created by the San Diego Super Computing Center as part of the development of a national HIS, and available at http://water.sdsc.edu/ODDataloader/ and discussed further in Section 3.4. The OD Data Loader, in its current preliminary prototype design, is limited to a specific configuration of fields. While very useful for many data sets, it would be best to load the entire SWQM database into ODM, not just the fields allowed by OD Data Loader. Thus, each applicable field will be mapped from

SWQM to ODM, and a SSIS script written to populate ODM in SQL Server. The SSIS script used is included in Appendix D.4.

4.4.2. Description of SWQM Database

Before a field map from SWQM to ODM can be completed, each data model must be understood. A complete description of the ODM Data Model can be found in the Working Design Specifications Document referenced above. A complete description of the TCEQ SWQM program can be found in *TCEQ*, 2006. A brief description of the database structure and contents is provided below.

The SWQM database consists of four tables: Event, Result, Stations and Parameter. The first two tables can be accessed as a pipe delimited text file by year and river segment from the TCEQ Sampling Data Query website,

http://www.tceq.state.tx.us/compliance/monitoring/crp/data/samplequery.html

(see Figure 4.6).

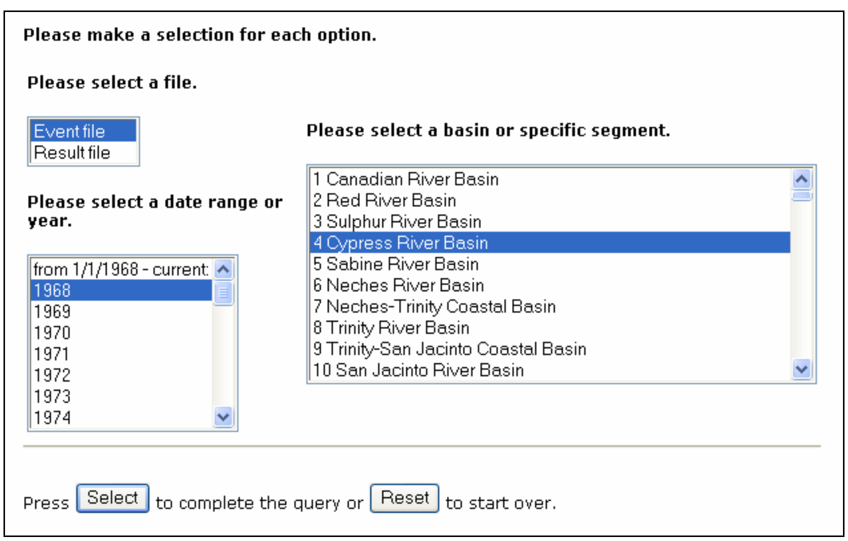


Figure 4.6 SWQM Events and Results File Access

The Stations table can be accessed at the Sampling Stations website, http://www.tceq.state.tx.us/compliance/monitoring/crp/data/station.html.

This is also in a pipe delimited text file format.

The Parameter table can be accessed at the Monitoring Parameter Descriptions website,

http://www.tceq.state.tx.us/compliance/monitoring/crp/data/storet.html.

The parameters and codes used are similar to those used by the EPA STORET system.

If converted to a Microsoft Access Database, these four tables would look like the diagram in Figure 4.7.

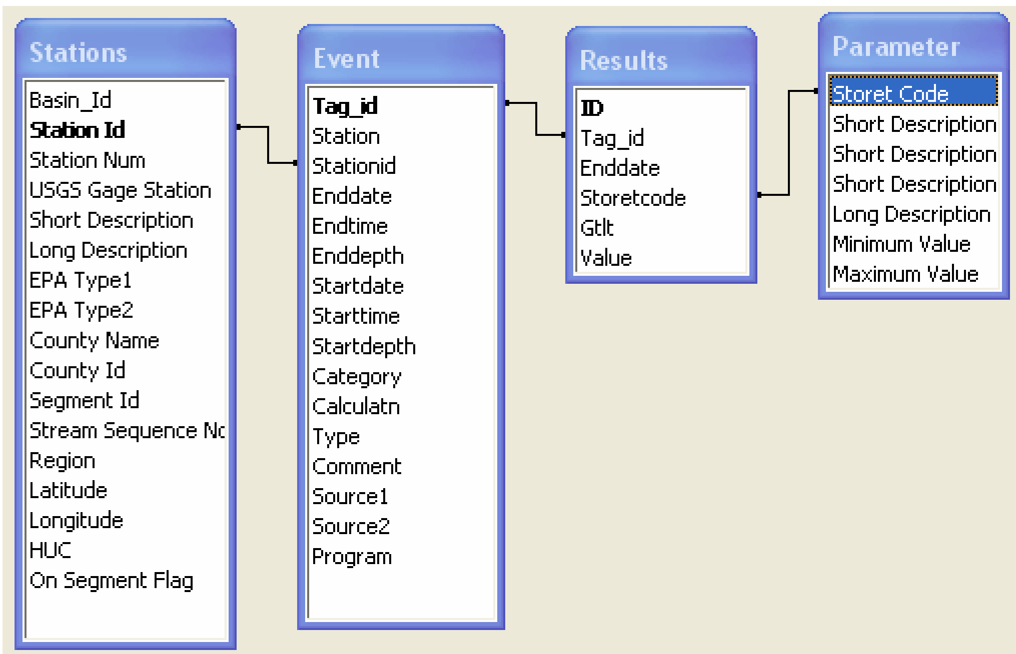


Figure 4.7 SWQM Tables Database Diagram

See Appendix D.1 for a list of fields and their descriptions. See Appendix D.2 for examples of each of the tables.

When water quality data is gathered for SWQM, information for both the Events and Results table is created. A record in the Events table describes a water quality sampling event. Each event has a location (station), date, time, source and other associated parameters. For instance, a private researcher may go out to a predetermined site (already registered with a Station Id) and deploy a multi-sensor device at a depth of 0.4 m below the surface of the water. The start and end date and time would be recorded. The private researcher may also make field notes, describing the weather, sampling conditions, or anything else that may be pertinent to the sample. The multi-sensor instrument would then measure a series of parameters. The results of these measurements would be recorded in the Results table. Thus, each event could have numerous associated results records. The information in the Event and Results table is linked via the Tag_id. Collectively, this information describes each record in the time series.

In general, the Stations table in SWQM is mapped to the Sites table in ODM, Parameter to Variables, and the combination of Event and Results to Values. The mapping of specific fields is described below.

4.4.3. Methods

The following methods were used in moving the TRACS SWQM data to an ODM 4.0 database.

4.4.3.1. Field Mapping

After careful analysis of each data model, the following field maps were developed (Table 4.1 thru Table 4.4). A few of the links between SWQM and ODM have multiple options. Possible solutions are discussed in Section 4.4.3.2. The final solution has been included in the following series of tables. Links that have multiple options, or otherwise require further discussion are highlighted with a gray background. This map is ordered by the SWQM table and field order. Additional information regarding field mapping can be found in Appendices C.3 and C.4.

Table 4.1 SWQM Events to ODM 4 Field Mapping

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
Event	Enddate	Values	DateTime	Add to Event:Endtime
Event	Endtime	Values	DateTime	Add to Event:Enddate
Event	Enddepth	Values	OffsetValue	
Event	Stationid	Values	SiteCode	SWQM Internal Identifier.
Event	Tag_id	Group-Descriptions	Group-Description	
Event	Category	Samples	SampleType	
Event	Type	Samples	SampleType	
Event	Comment	Qualifiers	QualifierDescription	
Event	Source1	Sources	Organization	
Event	Source2	Sources	Organization	
Event	Program	Sources	Organization	

Table 4.2 SWQM Parameters to ODM 4 Field Mapping

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
Parameter	Storet Code	Variables	VariableCode	
Parameter	Long Description	Variables	VariableName	

Table 4.3 SWQM Results to ODM 4 Field Mapping

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
Result	Gtlt	Values	CensorCode	If '<', lt; If >, 'gt'
Result	Value	Values	Value	
Result	Storetcode	Variables	VariableCode	

Table 4.4 SWQM Stations to ODM 4 Field Mapping

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
Stations	County Name	Sites	County	
Stations	Latitude	Sites	Latitude	
Stations	Longitude	Sites	Longitude	
Stations	Station Id	Sites	SiteCode	
Stations	Long Description	Sites	SiteName	
Stations	HUC	Sites	Comments	"HUC 8 = "
Stations	EPA Type1	Sites	Comments	"EPA Type1 = "
Stations	EPA Type2	Sites	Comments	"EPA Type2 = "

The following are fields that do not exist within SWQM, but have a common value for the entire database, and can thus be defined universally for this dataset.

Table 4.5 ODM 4 fields with common values

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
		OffsetTypes	OffsetUnitsID	52
		OffsetTypes	OffsetDescription	Depth below water surface level
		QualityControlLevels	QualityControlLevel	2
		QualityControlLevels	Definition	This data has been subjected to a limited quality control check.
		QualityControlLevels	Explanation	This data has been visually quality control checked. No systematic QC algorithms have been performed.
		Sites	State	Texas
		Sites	LatLongDatumID	2
		Sources	SourceDescription	Text file retrieved from TCEQ TRACS SWQM program, with data originally from numerous public and private monitoring organizations.
		Sources	SourceLink	http://www.tceq.state.tx.us/compliance/monitoring/crp/data/samplequery.html
		Sources	Email	crp@tceq.state.tx.us
		Sources	Address	TCEQ, Contact Name, Mail Code, P.O. Box 13087
		Sources	City	Austin
		Sources	State	TX
		Sources	ZipCode	78711-3087
		Values	UTCOffset	-6
		SpatialReferences	SpatialReferenceID	2
		SpatialReferences	SRSID	4269
		SpatialReferences	SRSName	NAD83
		SpatialReferences	IsGeographic	TRUE
		Values	UTCDateTime	=Values:DateTime – 6
		Values	QualityControlLevel	2

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
		Variables	SampleMedium	Surface Water
		Variables	ValueType	Field Observation
		Variables	GeneralCategory	Water Quality

4.4.3.2.Field Mapping Issues

The following are fields from SWQM that can be loaded into ODM using multiple methods, or require additional explanation. The method used in the SSIS package has been included in the field mapping tables above. Additional discussion and alternative methods are provided in this section.

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
Event	Enddate	Values	DateTime	Add to Event:Endtime
Event	Endtime	Values	DateTime	Add to Event:Enddate

The ODM4 DateTime field contains both Date and Time. SWQM stores Date and Time in separate fields. Thus, the two fields, EndDate and EndTime, should be concatenated in the format MM/DD/YYYY HH:MM:SS.

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field
Event	Tag_id		

The Tag_id in the SWQM database system is an essential link between the Results and the Events table. It also provides a permanent reference link to the original data, and describes a group of results that were collected together. An ODM table filled with SWQM data without a Tag_id loses a critical link to the original data. If a user of the ODM would need to track the source of an actual value, they would need the Tag_id to do so.

Recommended Solution:

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field
Event	Tag_id	GroupDescriptions	GroupDescription

- Each result record (Values in ODM) is linked to its associated Tag_id via the Groups and Group Descriptions table. The actual Tag_id would be recorded in the GroupDescription field in the GroupDescriptions table of ODM, making each group synonymous with a sampling event. Each group description is "Tag_id = xxxxx". Each Tag_id would be associated with the values collected as part of that specific event through the Groups table, matching GroupID and ValueID. Thus, a permanent link to the Events table in SWQM is maintained.

Alternative Solutions:

- Add the “EventDescriptions” table as a new entity to ODM. All the data in the Events table could then be migrated to SWQM with ease. The EventDescriptions table would be linked to the Values table through an intermediate table. See the Figure 4.8 for an example of how this might look.

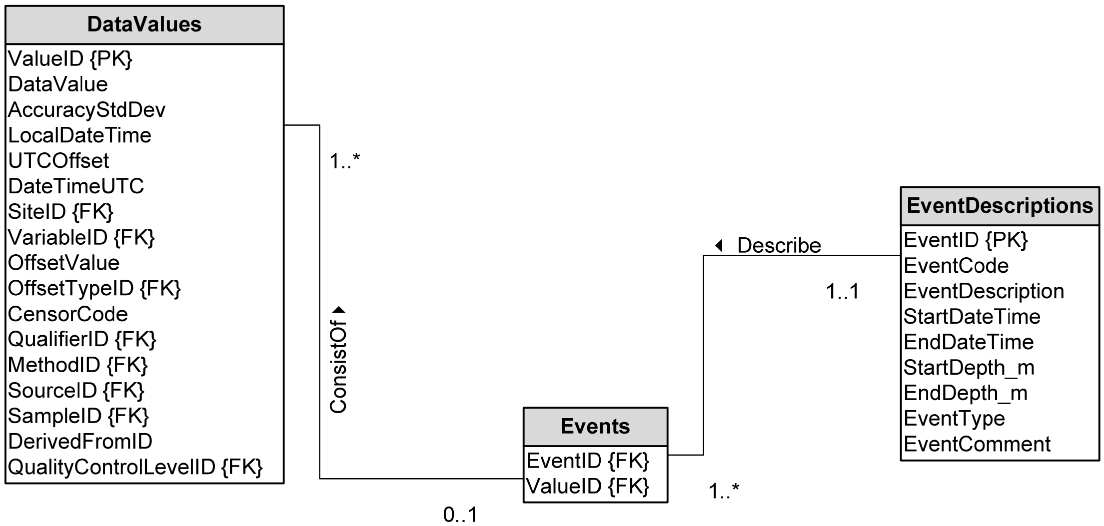


Figure 4.8 Event:TagID Inclusion Alternative

Figure courtesy of Jeff Horsburgh, Utah State University.

- Create a "ValueCode" field in the Values table of ODM. This is similar to the VariableCode and SiteCode fields which link records in the ODM table to records in the original database.

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field
Event	Category		
Event	Type		

The Event Category describes an event that is collected as a time composite (T), spatial composite (S), both (B) or a flow weighted composite (F). Some results reported in the Results table are actually the average of multiple "grabs" over a period of time or space. Each of these composite samples is further qualified in the Type field with "CN" for continuous composite, "##" to describe the number of grabs that make up the composite or "GB" when the number of grabs is not known.

Recommended Solution

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field
Event	Category	Samples	SampleType
Event	Type	Samples	SampleType

- Create a unique entry in the Samples table of the ODM for each combination of Category and Type. After a brief scan of SWQM data, it appears that a limited number of these combinations exist. Each time a Result is collected as part of an event that has a Type and Category, the corresponding record in the Values table would have a SampleID corresponding to the unique record in the Samples table of that specific Category and Type.

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field
Event	Comment		

Comments describing the sampling event are recorded in the Comment field of the Event table. The contents of the Comment field are usually similar to field notes. These do not necessarily use a controlled vocabulary, nor are they uniform throughout the database. See Appendix D.2.A for a sample Events table with example comments.

Some of these comments would best fit into the ODM as a separate Value with Categorical Variable records. However, it is critical for the understanding of the data that the link between 'real' value and categorical value remain intact. Less than thorough use of the data in ODM could miss some of these links between 'real' and categorical values, leading to erroneous data interpretation. Even if such a strategy were used, it could be extremely tedious to parse out all the possible categorical values from the Comment field.

Recommended Solution

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field
Event	Comment	Qualifiers	QualifierDescription

- Each comment from an event is entered as a record in the Qualifiers table in the ODM, under the QualifierDescription field. The specific records from each sampling event are then associated with the event comments via the QualifierID field in the Values table of ODM. This has the potential to create a large Qualifiers table, but seems to be the best solution to this issue.

Alternative Solution:

- If the Events table is added as a separate entity to ODM, as discussed above, Event comments would automatically be included. However, any query applications built on the standard ODM schema would have to be adjusted to access this additional table.

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field
Event	Source1		
Event	Source2		
Event	Program		

The Source1, Source2 and Program fields further trace the origins of the SWQM data. The TCEQ is a secondary source for some of the data provided in SWQM. Other primary agencies, individuals, or groups have submitted data to be included in SWQM.

Recommended Solution

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field
Event	Source1	Sources	Organization
Event	Source2	Sources	Organization
Event	Program	Sources	Organization

- Information for Source1, Source2, and Program are all combined, and placed in the Organization field of the Sources table in ODM. Thus, the original organization information is kept intact. The remaining Sources fields will be filled with information about the TCEQ SWQM program.

Alternative Solution

- If the Events table is added as a separate entity to ODM, as discussed above, Event Source1, Source2 and Program would automatically be included. However, any query applications built on the standard ODM schema would have to be adjusted to access this additional table.

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
Result	Gtlt	Values	CensorCode	If <, 'lt'; >, 'gt'

The Gtlt field from the Result table in the SWQM is equivalent to the CensorCode in the Values table in ODM. If the value of CensorCode should be "lt" if the value of Gtlt is "<", and the value of CensorCode should be "gt" if the value of Gtlt is ">".

Additionally, the table CensorCodeCV (controlled vocabulary) should contain entries explaining that "lt" is "less than" and "gt" is "greater than." This can be imported from an existing CensorCodeCV table, or created manually.

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
Stations	HUC	Sites	Comments	"HUC 8 = xxxxxxxx"

Like the Tag_id issue, this has been discussed by the CUAHSI community. It has been agreed that a new "HUC" field should not be added to the Sites table. As additional information, it will be included in the Comments field of the Sites table. Thus, the Comment in each Sites record should include "HUC 8 = xxxxxxxx" where xxxxxxxx is the record imported from SWQM.

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field
Stations	EPA Type1		
Stations	EPA Type2		

The EPA Type1 and EPA Type2 fields in the Stations table further describe the sampling location. A typical entry for EPA Type1 would be "RESERV" indicating that the sampling location is in a reservoir. A typical entry for EPA Type2 would be "AMBNT" indicating that the sampling location is exposed to ambient conditions.

Recommended Solution

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	
Stations	EPA Type1	Sites	Comments	"EPA Type1 = "
Stations	EPA Type2	Sites	Comments	"EPA Type2 = "

- Each of these fields could be included with the HUC field from the Stations table in the "Comments" field in the Sites table of ODM4. However, this begins to clutter that single field.

Alternative Solutions:

- Add a SiteType field in the Sites table of ODM, and a SiteTypeCV table that describes possible types. This is currently being discussed by the creators of ODM.

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field
		Units	UnitsID
		Units	UnitsName

The Units table in the ODM is meant to further describe the units of each Variable, which in turn describes each value. Thus, a single variable (water temperature) can be measured in terms of multiple units (degrees Celsius, degrees Fahrenheit). The parameter descriptions from SWQM (which closely match those from the legacy STORET parameters list) include units with the variable description. Thus, the instead of a single variable, multiple variables with units are listed in the parameters table (water temperature in degrees Celsius, water temperature in degrees Fahrenheit). This complete parameter description is included in the Parameter:LongDescription field from SWQM, which is imported into the ODM Variables:VariablesName field. Thus, no further description of the units is necessary. Additionally, parsing out the units from each of these descriptions would be extremely difficult and time consuming.

Recommended Solution:

- Not include any information in the Units table within ODM. The units description within the Variables:VariablesName field is sufficient.

Alternative Solution:

- Separate the Variable from the Units within the SWQM table, and include both in their respective separate table within the ODM. Because the SWQM Parameters are based on the EPA legacy STORET parameters, the majority of this work may have already been completed by other groups using the STORET parameters.

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field
		Sites	Elevation_m
		Sites	VerticalDatum

There is no elevation information describing any of the Stations within the SWQM database. However, given a horizontal location and a Digital Elevation Model (DEM), the vertical location of the site could be determined. This assumes that the elevation of the sampling site occurs at the ground surface elevation.

Recommended Solution:

- Intersect the site locations with a DEM, and record the corresponding elevation in the Sites:Elevation_m field. The datum used to describe the DEM would be the Sites:VerticalDatum.

Alternative Solution:

- Do not include any Elevation information in the ODM

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
		Values	UTCDateTime	=Values:DateTime - 6

The UTCDateTime field describes the time of a sample in the Coordinated Universal Time (UTC). The value of UTCDateTime needs to be calculated during the transfer of information from SWQM to ODM. UTCDateTime is the value of DateTime minus six (-6) hours during standard time, and minus five (-5) hours during daylight savings time. A method of distinguishing between the two still needs to be developed.

4.4.3.3. Order of Operations

The order of loading data into the ODM is critical to the success of the data transfer. In general, the tables need to be loaded with data from the outside towards the middle of the ODM table diagram (see Figure 2.1). The outer tables are support tables that are not dependent on values in the critical inner tables. For instance, the Sites table is dependent on values in the SpatialReferences table for values in the fields LatLongDatumID and LocalProjectionID. Thus, the SpatialReferences table needs to be filled before Sites can be filled. The order of data migration of any data source into the ODM can be generalized, although a few changes may be required for some specific sources. For instance, the use of the DerivedFrom table within the ODM creates additional complications when migrating data to the Values table due to the interdependencies between the Values and DerivedFrom tables. Because the DerivedFrom table is not used with the SWQM data source, it is not included in the following order of operations.

The order of operations for SWQM to ODM is described in Table 4.6 below. The order of data migration within a single group is not critical. It is only necessary to load all tables in group 1 before those in group 2.

Table 4.6 ODM Data Migration Order of Operations

Group	Table	Critical Dependent Tables
1	GroupDescriptions	
	ISOMetadata	
	LabMethods	
	Methods	
	Qualifiers	
	QualityControlLevels	
	SpatialReferences	
	Units	
2	OffsetTypes	Units
	Samples	LabMethods
	Sites	SpatialReferences
	Sources	ISOMetadata
	Variables	Units
3	Categories	Variables
4	Values	Samples, Sources, Methods, Variables, Sites, OffsetTypes, Qualifiers, QualityControlLevels
5	Groups	GroupDescriptions, Values
	SeriesCatalog	Sites, Variables, Units, Values

4.4.4. Conclusions

The creation of TxHIS is an exciting development in information availability in Texas. The utility of such a system will be dependent on three things: the quantity, quality, and accessibility of the data. The addition of the SWQM database to the TxHIS is a test of the ability to provide both quantity and quality data from an existing data source.

This section provides a map from which SWQM data was moved to the ODM schema. Solutions for each of issues have been presented. In some cases, multiple solutions have been presented, with the final solution having been identified.

4.5. SWQM DATA ANALYSIS

One of the advantages of moving the SWQM data into a relational database such as the ODM is that SQL Queries can be written to analyze the data. The following analysis was performed to quantify the extent and range of SWQM data. A similar analysis could be performed using similar queries with any other dataset.

4.5.1. Introduction

On March 30, 2007, data from the SWQM portion of TRACS was downloaded from http://www.tceq.state.tx.us/compliance/monitoring/crp/data/samplequery.html.

Event and Results text files were downloaded for each of the 25 river and coastal basins in Texas. Table 4.7 lists these basins with the size of the corresponding Events and Results text files. Figure 4.9 shows the same data in a geospatial context.

Table 4.7 SWQM .txt file size by basin

Basin	Basin Name	Events Text File Size (MB)	Results Text File Size (MB)
1	Canadian River Basin	0.621	2.361
2	Red River Basin	1.423	7.141
3	Sulphur River Basin	0.503	2.828
4	Cypress River Basin	1.321	4.527
5	Sabine River Basin	2.819	15.760
6	Neches River Basin	2.931	11.439
7	Neches-Trinity Coastal Basin	0.663	2.817
8	Trinity River Basin	4.984	26.618
9	Trinity-San Jacinto Coastal Basin	0.088	0.332
10	San Jacinto River Basin	4.019	21.688
11	San Jacinto-Brazos Coastal Basin	8.590	3.995
12	Brazos River Basin	4.962	26.118
13	Brazos-Colorado Coastal Basin	0.196	0.934
14	Colorado River Basin	7.442	35.966
15	Colorado-Lavaca Coastal Basin	0.119	0.458
16	Lavaca River Basin	0.578	2.455
17	Lavaca-Guadalupe Coastal Basin	0.070	0.231
18	Guadalupe River Basin	1.567	6.035
19	San Antonio River Basin	2.053	12.640
20	San Antonio-Nueces Coastal Basin	0.109	0.643
21	Nueces River Basin	0.666	4.292
22	Nueces-Rio Grande Coastal Basin	0.325	2.084
23	Rio Grande River Basin	1.761	12.525
24	Bays and Estuaries	6.836	24.982
25	Gulf of Mexico	0.226	0.937
	Total Size	54.872	229.806
	Average	2.195	9.192
	Minimum Size	0.070	0.231
	Maximum Size	8.590	35.966

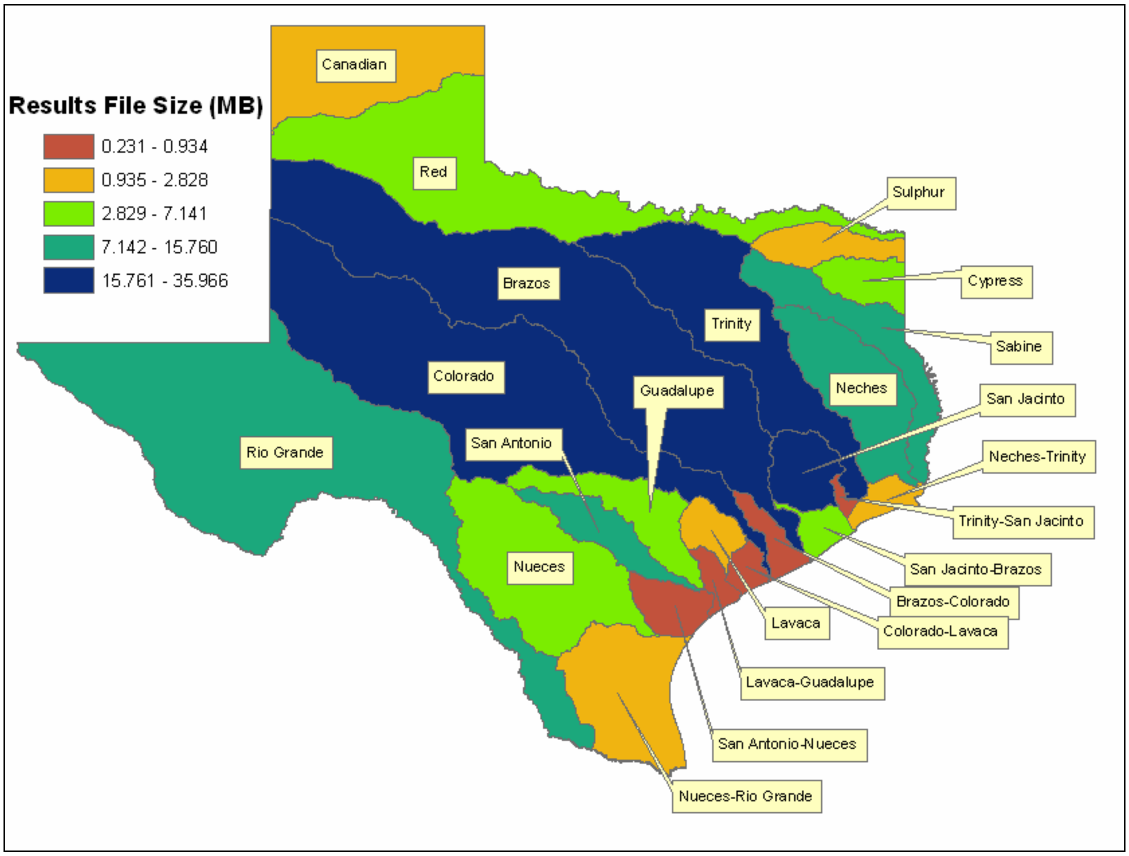


Figure 4.9 SWQM .txt file size by basin

After downloading the text files, the data was inserted into a Microsoft Access Database. Along with the Events and Results files, the Stations and Parameters files were also added, further describing the data. The MS Access Database (.mdb) with the complete SWQM dataset from 1/1/1968 to 3/30/2007 uses 928 MB of space. This same data has been migrated into the ODM data format on a SQL Server platform, described in section 4.4. The ODM/SQL version of the SWQM data uses nearly 10 GB of space.

Using a series of SQL queries, the information shown in Figure 4.10 through Figure 4.30, Table 4.8 and Table 4.9 about the entire SWQM dataset was gathered about SWQM data in Texas from the Access database.

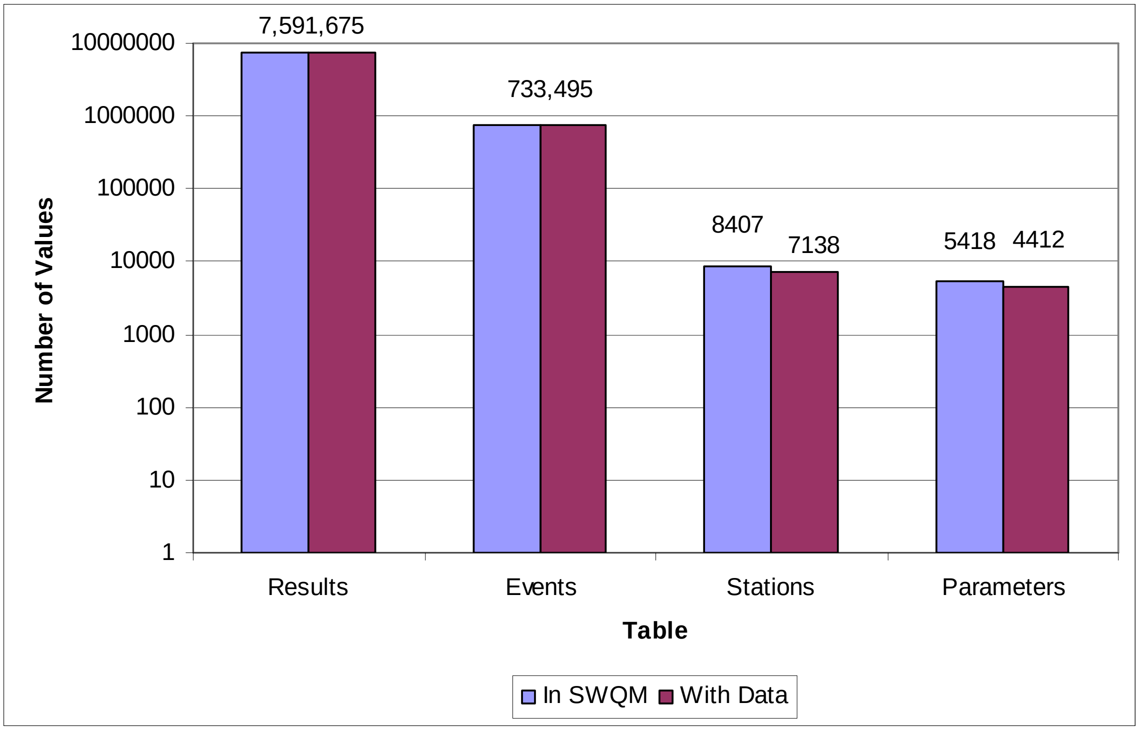


Figure 4.10 Number of SWQM Values by Table

The reason for the difference between "In SWQM" and "With Data" for the Stations table in Figure 4.10 is unclear. It appears that information was gathered for stations where data collection never actually occurred. It is possible that the list of stations was developed from a larger list of potential stations. The reason for the difference in the Parameters table is that the SWQM parameters are based off the EPA STORET codes. There are some EPA STORET codes that have never been measured in Texas, even though their description occurs in the Parameters table.

4.5.2. Events

The Events table describes each sampling event. An example of a sampling event is a scientist taking a bucket of water from a reservoir. That bucket of water can then be tested for multiple different parameters, yielding multiple results.

- The maximum number of records per event is 337
- 2589 events have more than 100 records per event
- 12,779 (1.7%) of the events have only 1 record per event
- 485,741 (66.2%) of the events have less than 10 records per event
- On average, each event has 10 records

4.5.3. Stations

Each location from which surface water quality data is gathered is considered a station. Some stations are locations with permanent instrumentation, and may gather data daily, hourly or even more frequently. These stations produce thousands of results. Other stations are locations from which a single sampling event occurred, and only a few results measured. Depending on the instrumentation available, hundreds and sometimes thousands of parameters can be measured at a single station. Depending on the station, this sampling may have occurred once (for a single sampling event), or periodically over a long time. For those stations with a large date range, sampling may have occurred regularly or irregularly. Data describing the statistics of stations can be found in Figure 4.11 through Figure 4.17.

- The maximum number of results for a single station is 48007 (StationID = 12302, right above Mansfield Dam on Lake Travis).

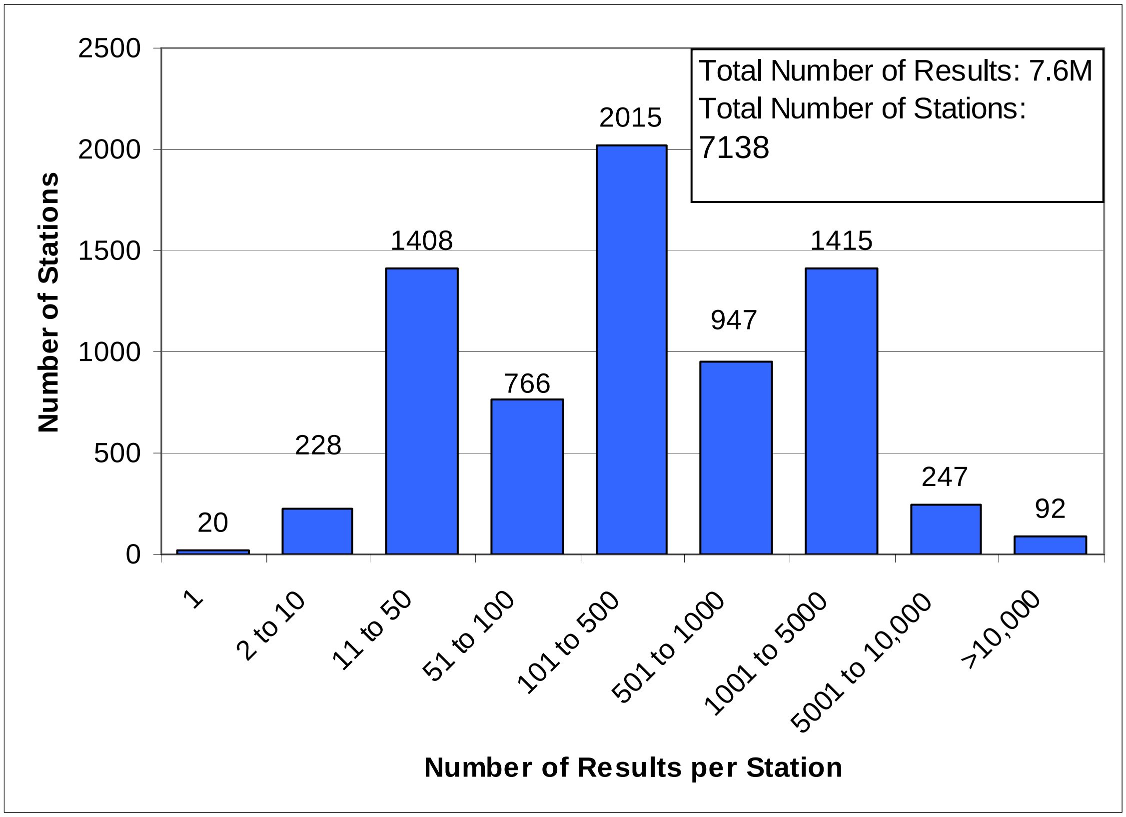


Figure 4.11 Station Results Frequency

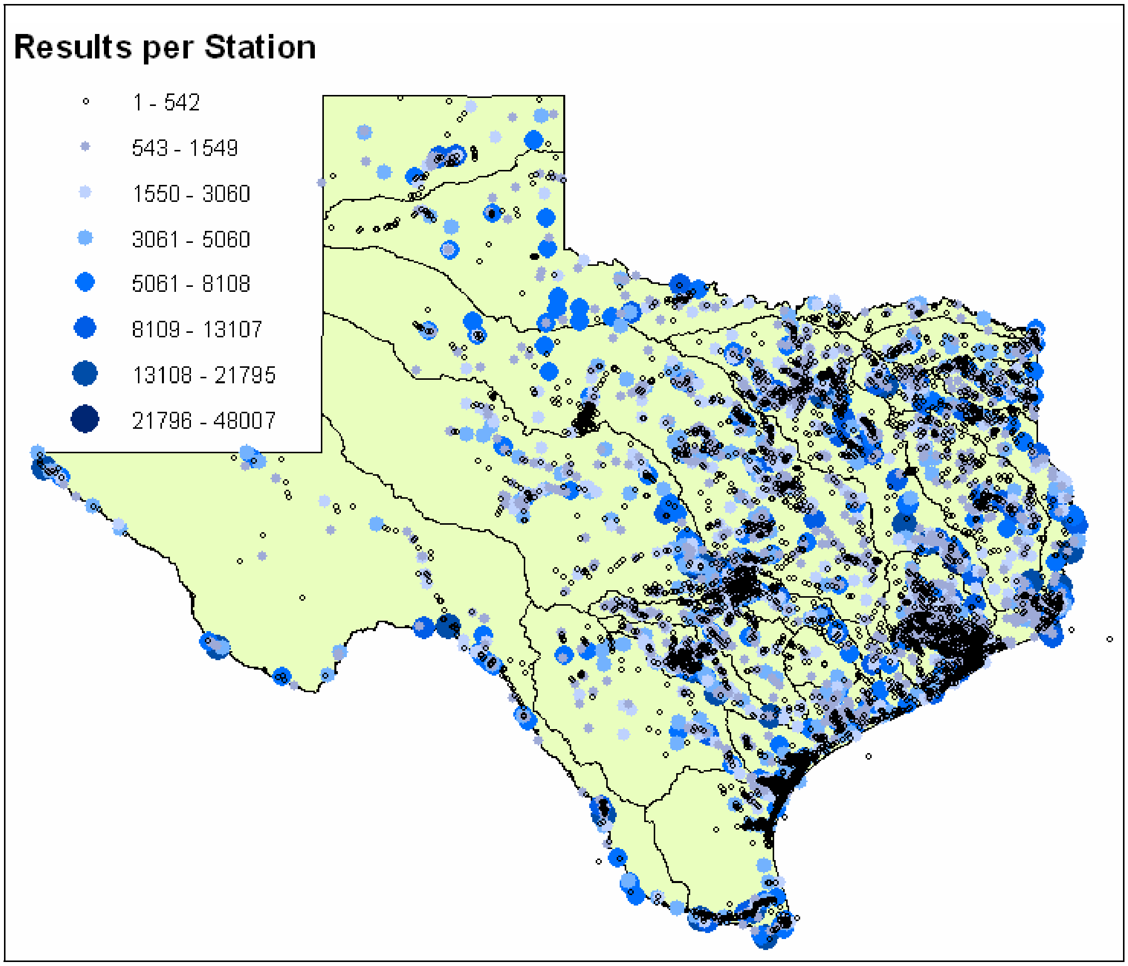


Figure 4.12 Results per Station

- The maximum number of parameters measured at a single station is 1517 (StationID = 11252)

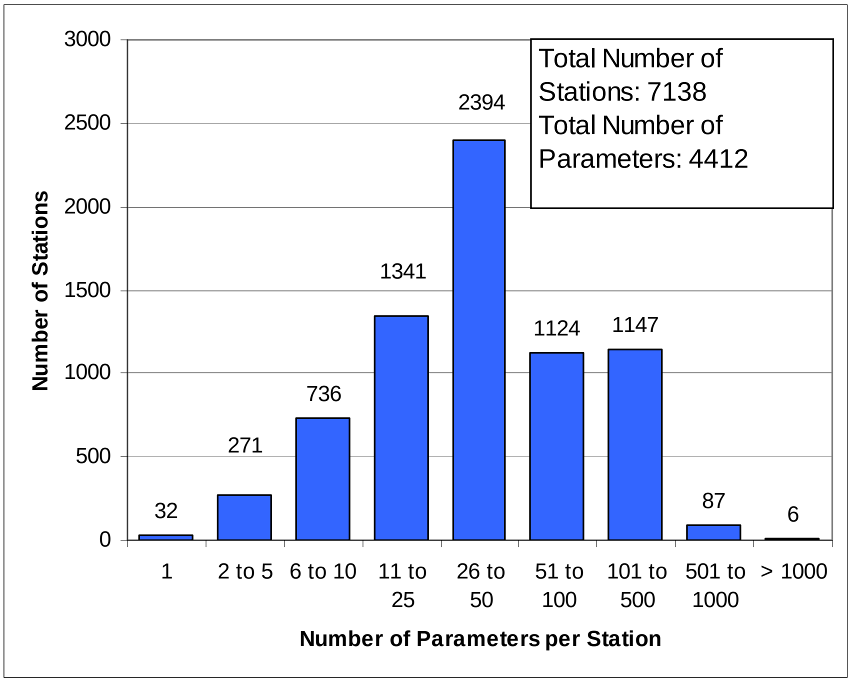


Figure 4.13 Station Parameter Frequency

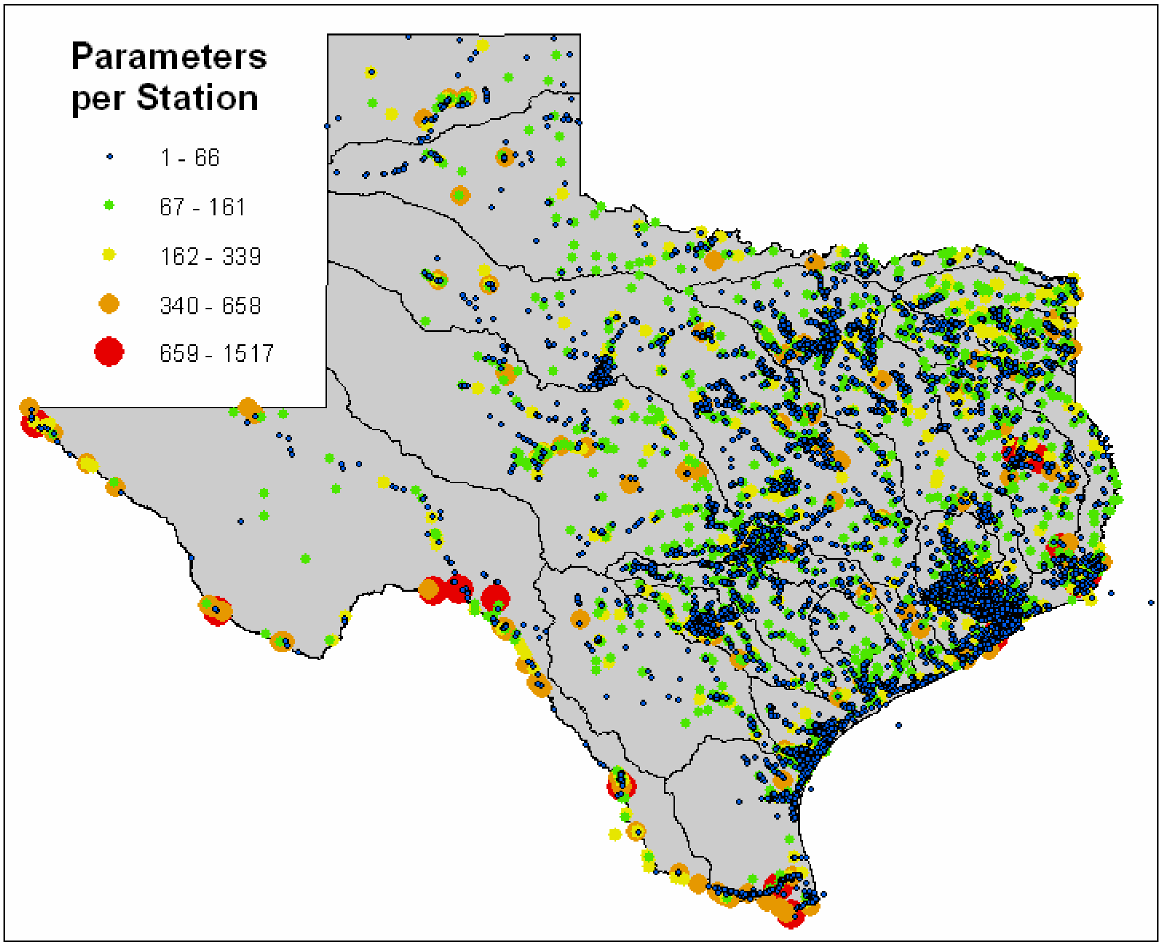


Figure 4.14 Parameters per Station

- The maximum number of events measured at a single station is 8660 (StationID = 12302)

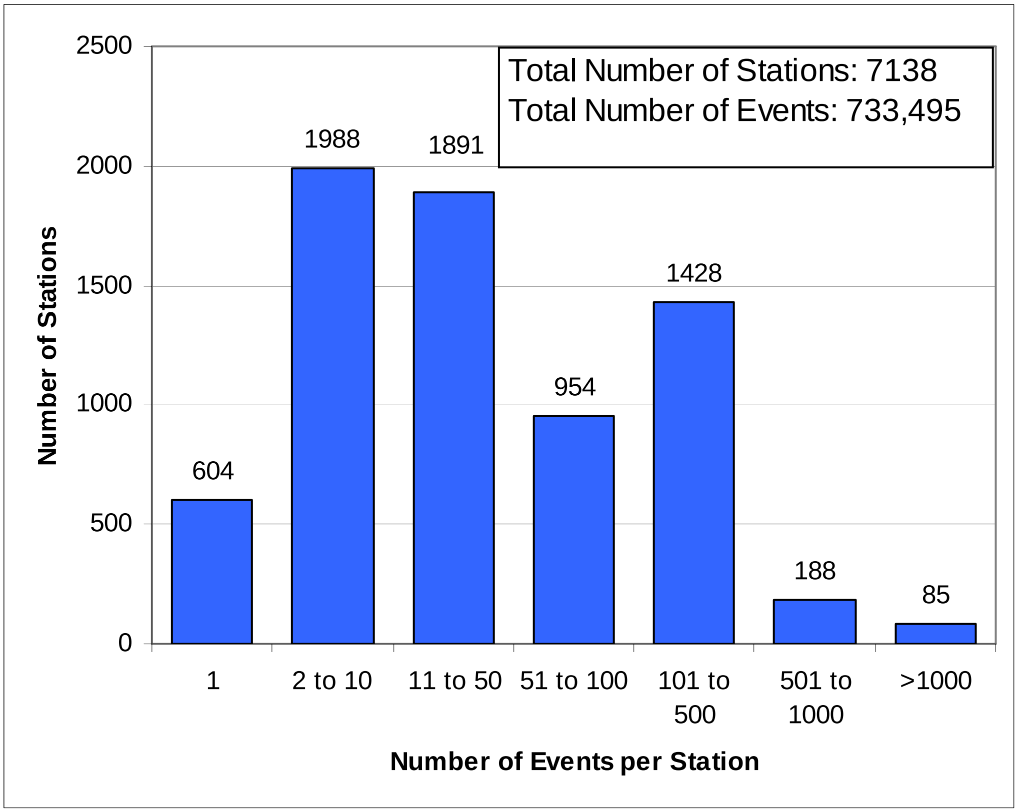


Figure 4.15 Station Event Frequency

- The date range of the events (and thus results) is from 2/4/68 to 8/14/06.

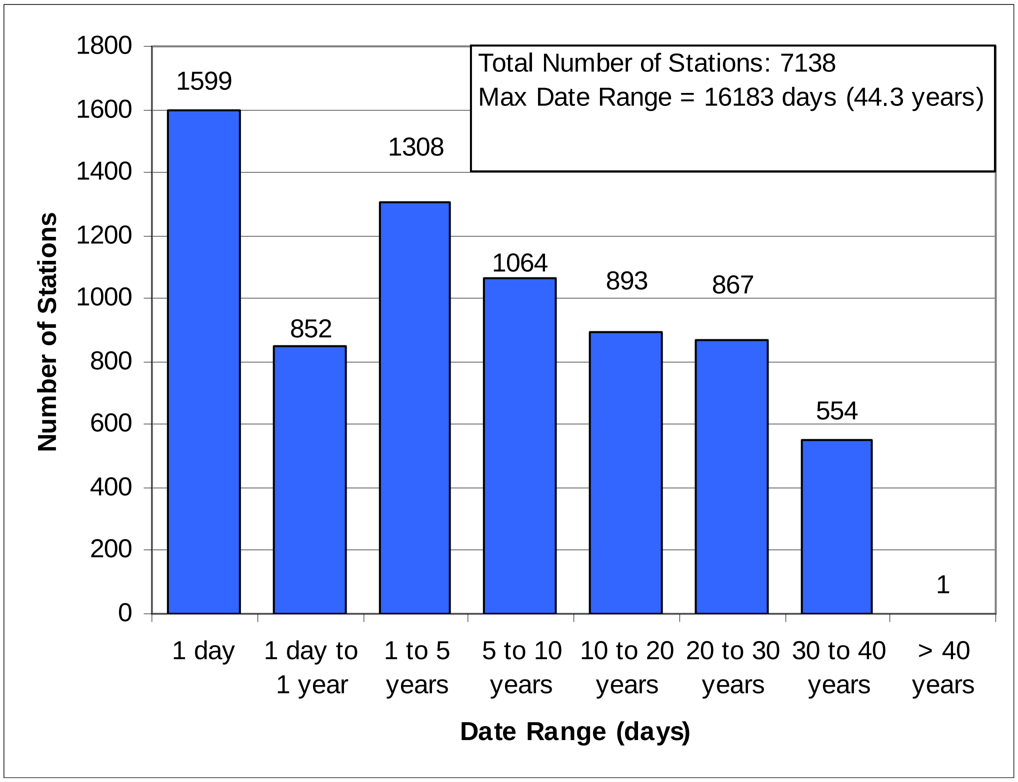


Figure 4.16 Station Date Range Frequency

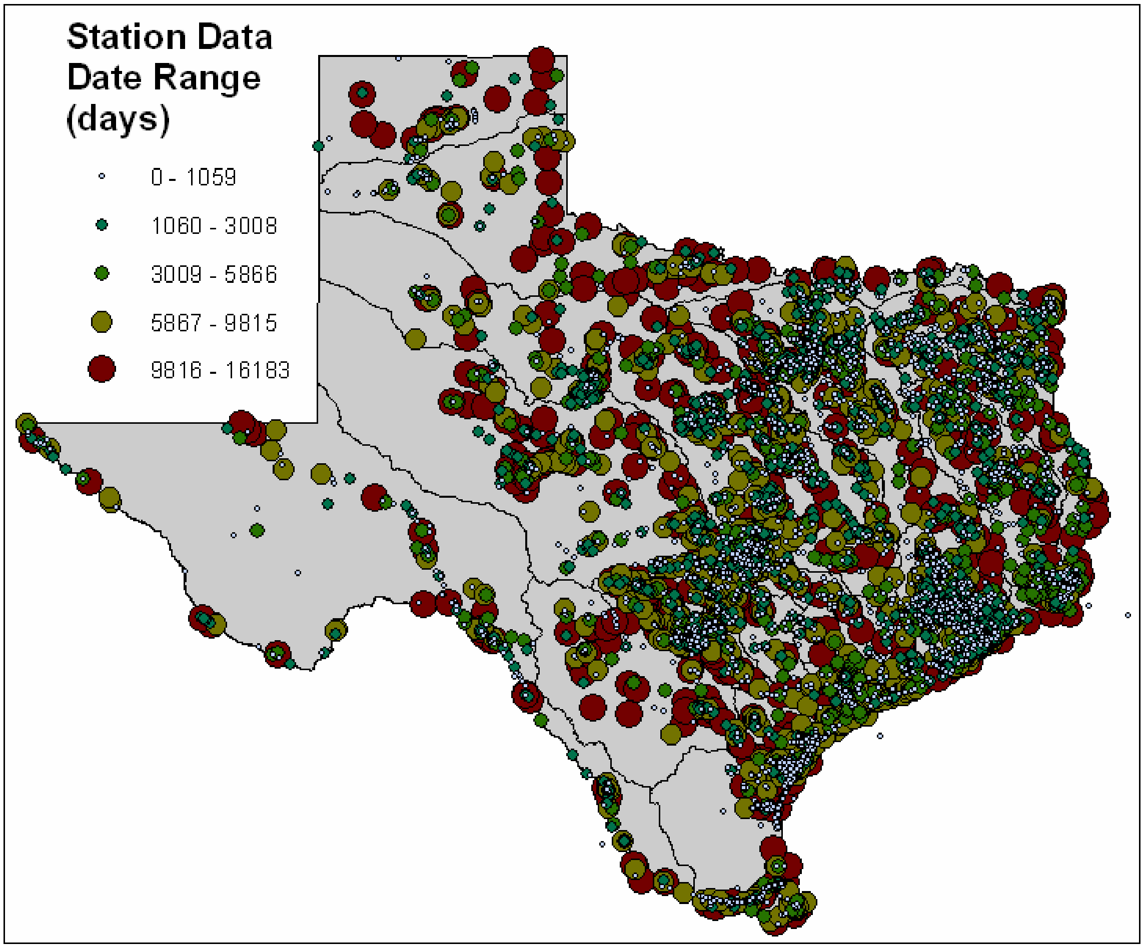


Figure 4.17 Station Data Date Range

4.5.4. Parameters

The parameters table further describes each value measured. Some commonly measured parameters are temperature, dissolved oxygen, pH and certain chemical concentrations such as Nitrogen and Phosphorous. The SWQM parameters table is based off the EPA legacy STORET codes. For this reason, some parameters in the list have no corresponding values. Data describing the statistics of the parameters can be found in Figure 4.18 through Figure 4.20, Table 4.8 and Table 4.9.

- The parameter with the most results is Water Temperature (degrees Celsius), with 592,383 values measured.

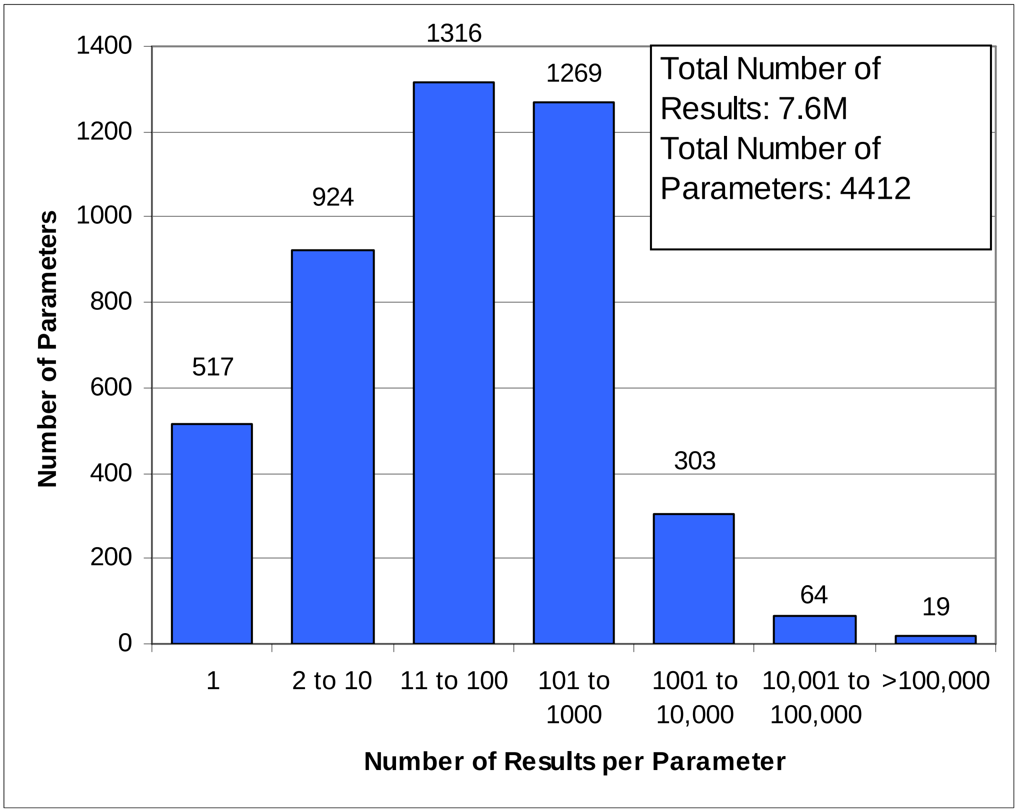


Figure 4.18 Parameter Result Frequency

Table 4.8 20 parameters with most records

Storet code	Long Description	ValueCount
00010	TEMPERATURE, WATER (DEGREES CENTIGRADE)	592,382
00300	OXYGEN, DISSOLVED (MG/L)	555,182
00400	PH (STANDARD UNITS)	517,462
00094	SPECIFIC CONDUCTANCE,FIELD (UMHOS/CM @ 25C)	475,349
00940	CHLORIDE (MG/L AS CL)	203,168
00945	SULFATE (MG/L AS SO4)	194,017
00530	RESIDUE, TOTAL NONFILTRABLE (MG/L)	179,647
00610	NITROGEN, AMMONIA, TOTAL (MG/L AS N)	166,895
00665	PHOSPHORUS, TOTAL, WET METHOD (MG/L AS P)	158,683
00011	TEMPERATURE, WATER (DEGREES FAHRENHEIT)	145,621
31616	FECAL COLIFORM,MEMBR FILTER,M-FC BROTH, #/100ML	142,285
70507	ORTHPHOSPHATE PHOSPHORUS,DISS,MG/L,FILTER >15MIN	131,931
00480	SALINITY - PARTS PER THOUSAND	123,478
00410	ALKALINITY, TOTAL (MG/L AS CACO3)	122,493
00680	CARBON, TOTAL ORGANIC, NPOC (TOC), MG/L	118,418
00620	NITRATE NITROGEN, TOTAL (MG/L AS N)	112,336
00095	SPECIFIC CONDUCTANCE (UMHOS/CM @ 25C)	108,731
00535	RESIDUE, VOLATILE NONFILTRABLE (MG/L)	108,083
32211	CHLOROPHYLL-A UG/L SPECTROPHOTOMETRIC ACID. METH	105,960
00301	OXYGEN, DISSOLVED (PERCENT OF SATURATION)	98,277

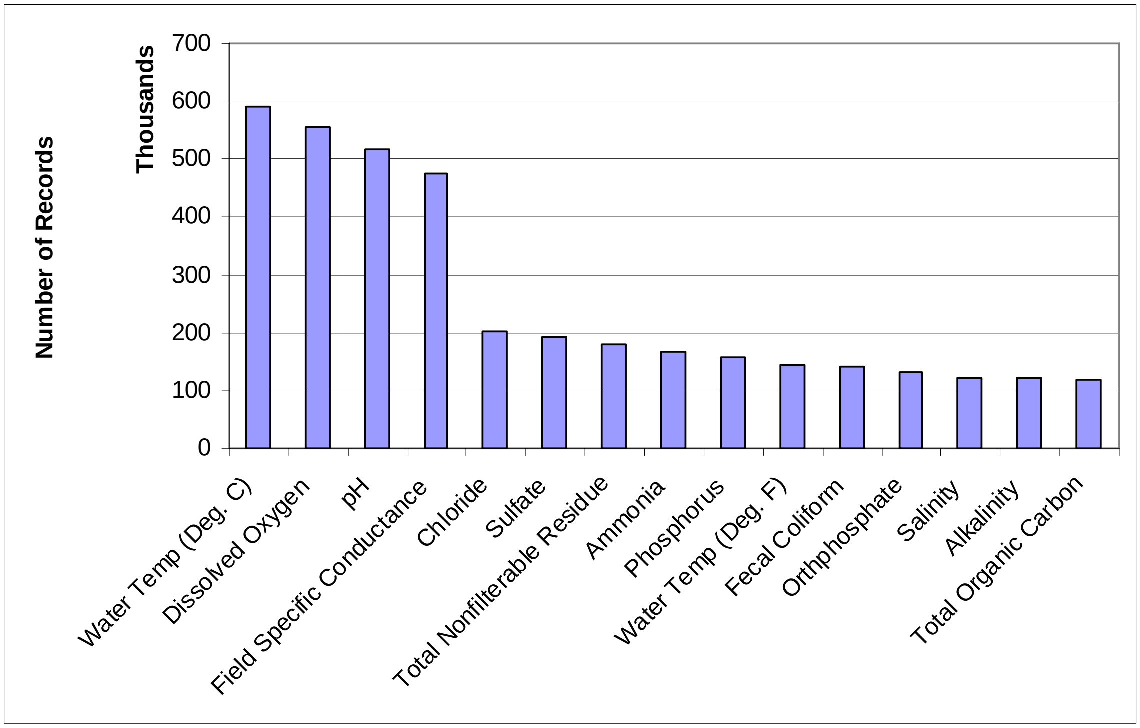


Figure 4.19 15 Parameters with most records

- The parameter that is measured at the most stations is also Water Temperature (degrees Celsius), which is measured at 6438 stations.

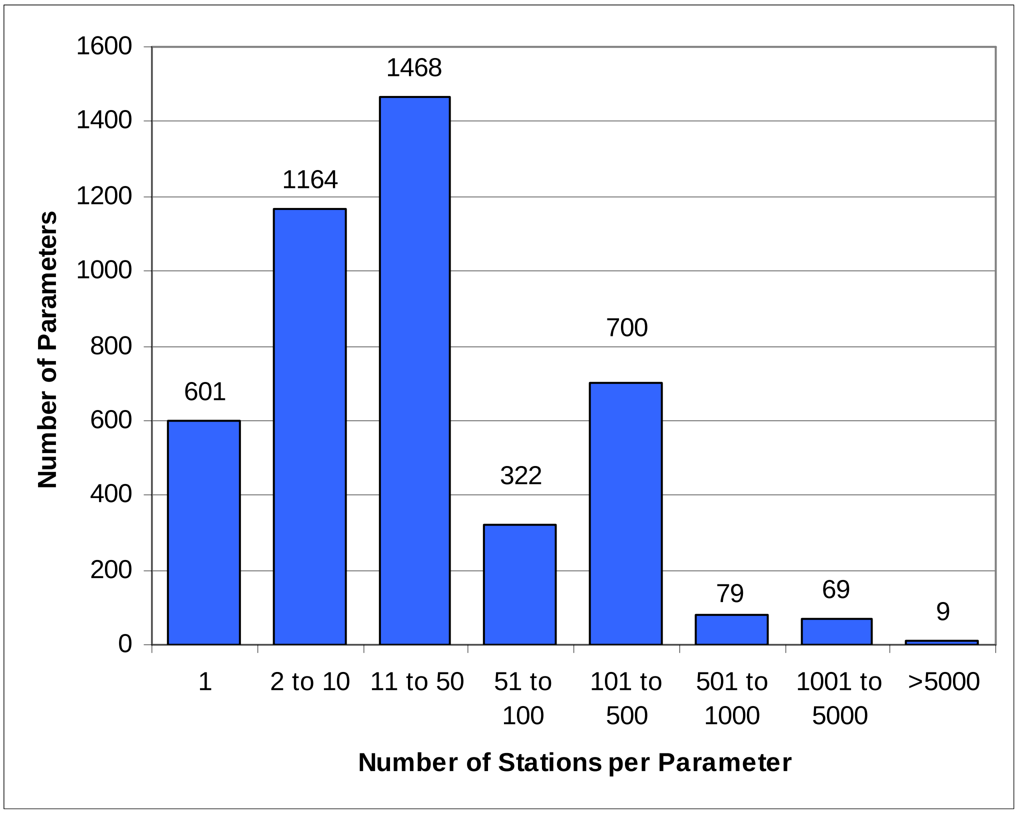


Figure 4.20 Parameter Station Frequency

Table 4.9 20 parameters measured at most number of sites

Storet Code	Long Description	Station Count	Value Count	Value Rank
00010	TEMPERATURE, WATER (DEGREES CENTIGRADE)	6438	592,382	1
00300	OXYGEN, DISSOLVED (MG/L)	6283	555,182	2
00400	PH (STANDARD UNITS)	5924	517,462	3
00094	SPECIFIC CONDUCTANCE,FIELD (UMHOS/CM @ 25C)	5590	475,349	4
00940	CHLORIDE (MG/L AS CL)	5327	203,168	5
00945	SULFATE (MG/L AS SO4)	5211	194,017	6
00530	RESIDUE, TOTAL NONFILTRABLE (MG/L)	5166	179,647	7
00610	NITROGEN, AMMONIA, TOTAL (MG/L AS N)	5072	166,895	8
00665	PHOSPHORUS, TOTAL, WET METHOD (MG/L AS P)	5062	158,683	9
70507	ORTHPHOSPHATE PHOSPHORUS,DISS,MG/L,FILTER >15MIN	4600	131,931	12
00410	ALKALINITY, TOTAL (MG/L AS CACO3)	4346	122,493	14
32211	CHLOROPHYLL-A UG/L SPECTROPHOTOMETRIC ACID. METH	4283	105,960	19
70300	RESIDUE,TOTAL FILTRABLE (DRIED AT 180C) (MG/L)	4184	86,231	23
00535	RESIDUE, VOLATILE NONFILTRABLE (MG/L)	4123	108,083	18
00620	NITRATE NITROGEN, TOTAL (MG/L AS N)	4100	112,336	16
00630	NITRITE PLUS NITRATE, TOTAL 1 DET. (MG/L AS N)	4058	85,818	24
32218	PHEOPHYTIN-A UG/L SPECTROPHOTOMETRIC ACID. METH.	4037	80,999	26
00625	NITROGEN, KJELDAHL, TOTAL (MG/L AS N)	4031	91,095	21
31616	FECAL COLIFORM,MEMBR FILTER,M-FC BROTH, #/100ML	3899	142,285	11
00680	CARBON, TOTAL ORGANIC, NPOC (TOC), MG/L	3627	118,418	15

4.5.5. Basin Date Range

The following is a temporal analysis of data collection by basin. Due to the increasing need for surface water quality data and the increasing availability of remote sensing equipment, the overall rate of collection of surface water quality data continues to increase. While this is true for the state of Texas in general, it is not necessarily true for each individual basin. For example, the Lavaca-Guadalupe Coastal Basin has had a declining number of water quality records for each decade since the 1970s (see Figure

4.27). For additional data describing the data collected by basin, see Figure 4.21 through Figure 4.30.

Note that the date range for SWQM data only goes from 2/4/68 to 8/14/06, so the 1960's and 2000's decades only includes data for less the 10 years of data displayed in the other decades.

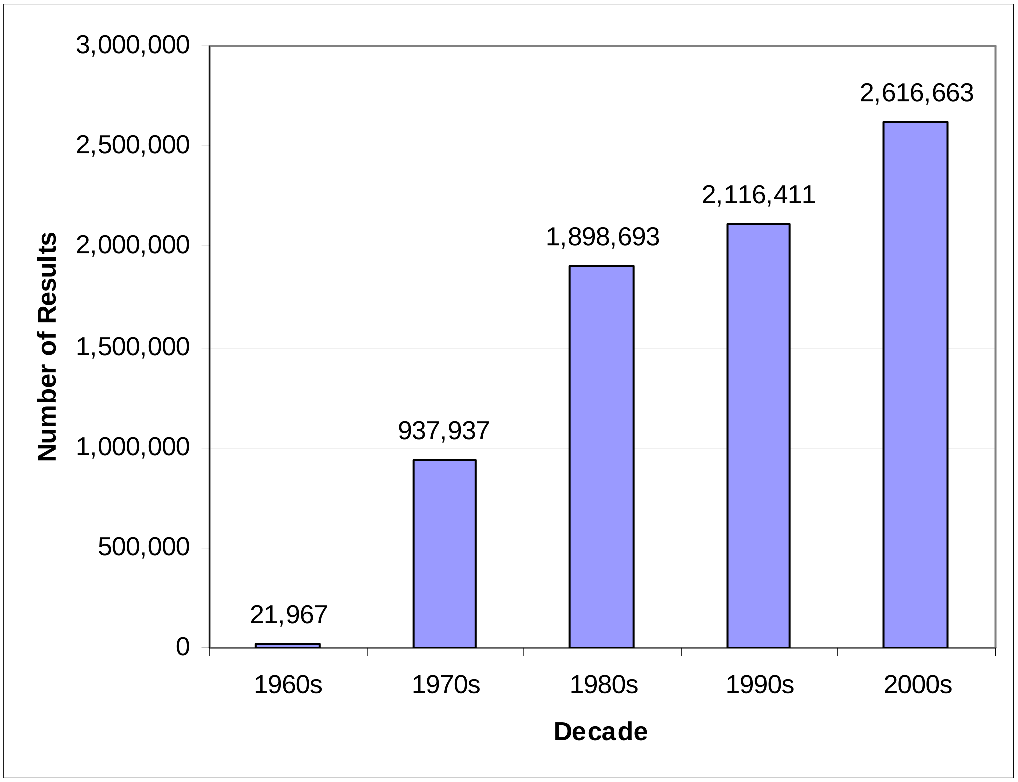


Figure 4.21 SWQM Results by Decade: Texas

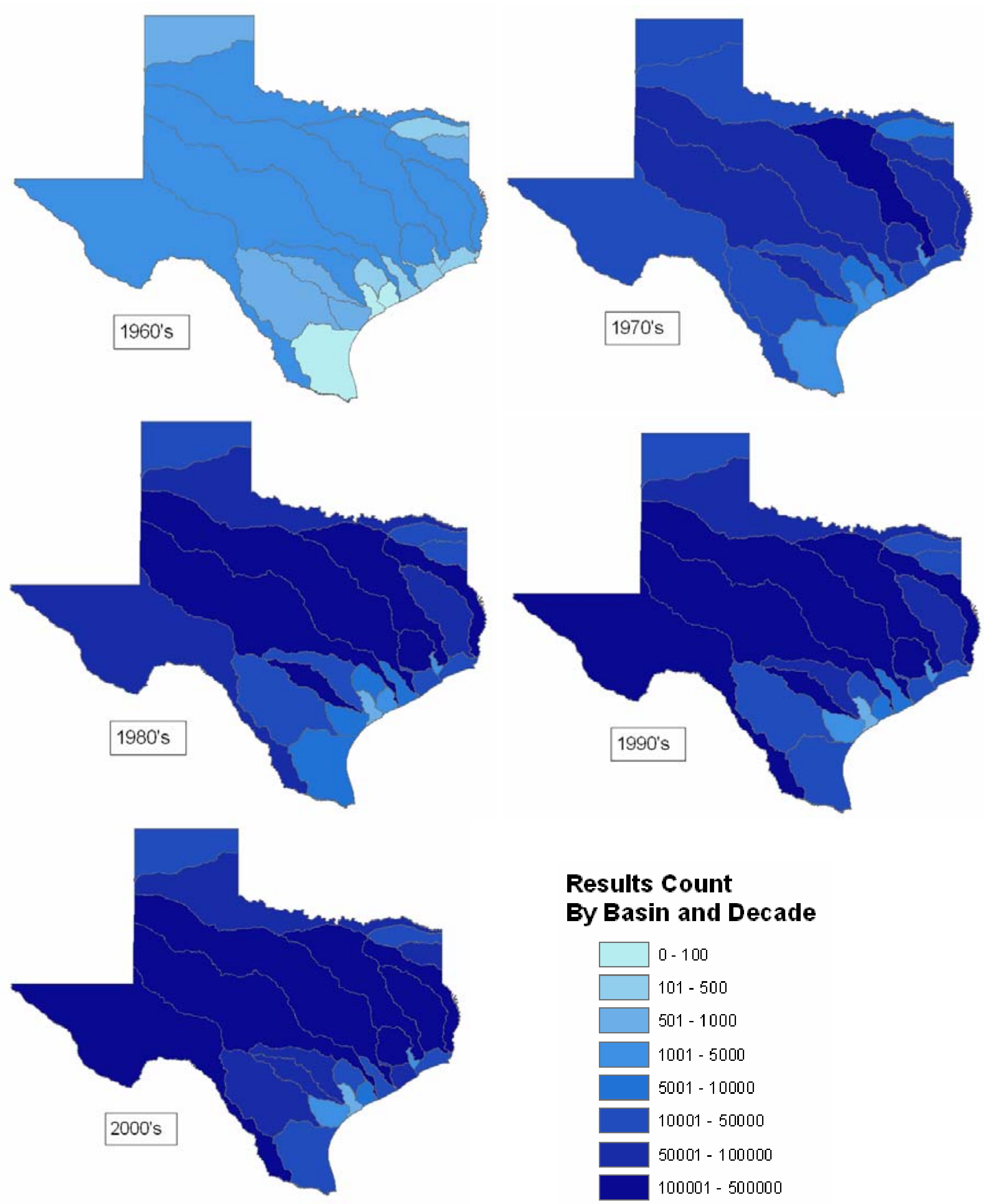


Figure 4.22 Results per Basin by Decade

The map of basins in Figure 4.23 is color coded to correspond with decadal graphs of water quality sampling for a selected series of basins, found in Figure 4.24 through Figure 4.30. These basins were selected for their size of sampling, as well for unique decadal patterns in sampling.

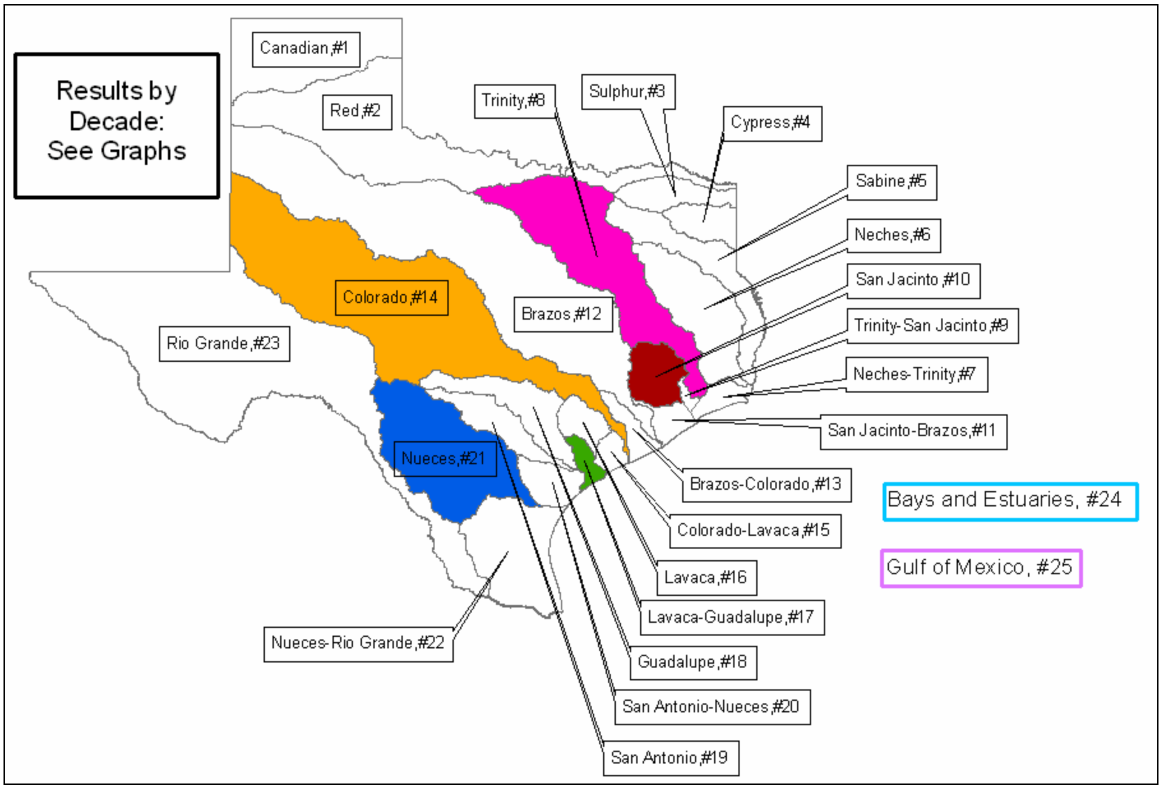


Figure 4.23 Results by Decade Map

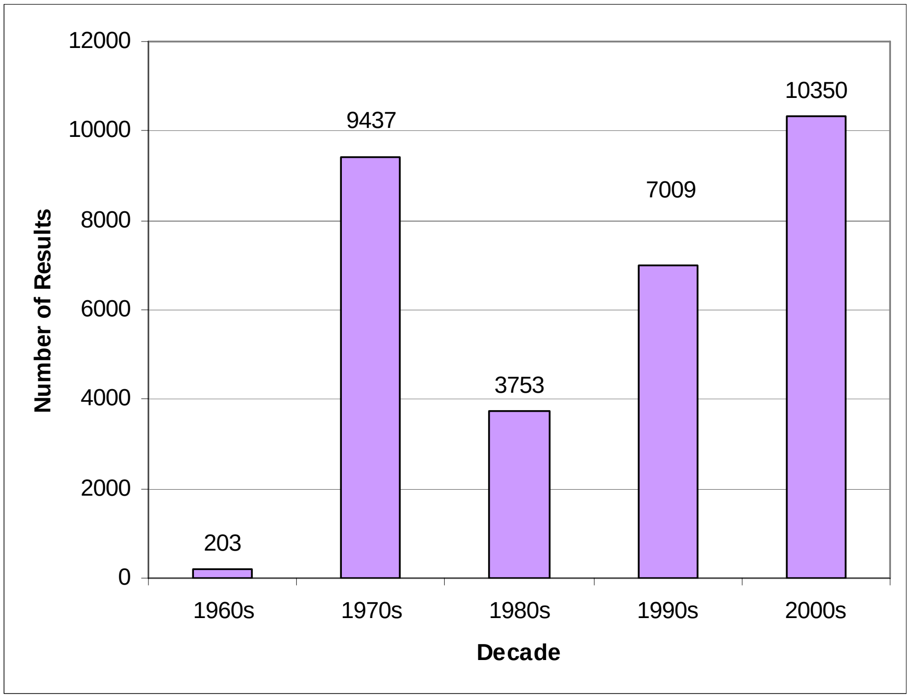


Figure 4.24 SWQM Results by Decade: Gulf of Mexico, #25

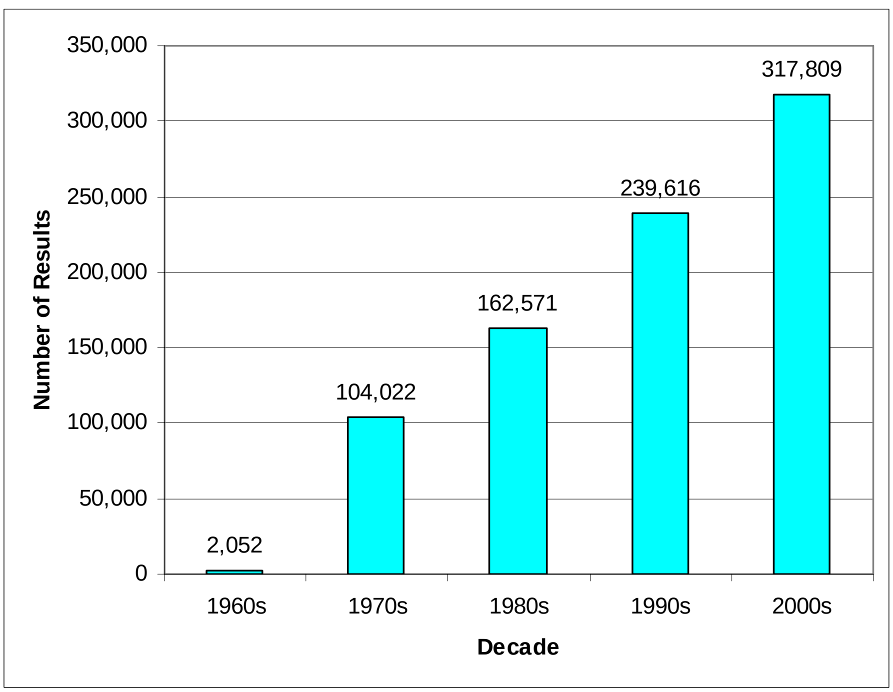


Figure 4.25 SWQM Results by Decade: Bays and Estuaries, #24

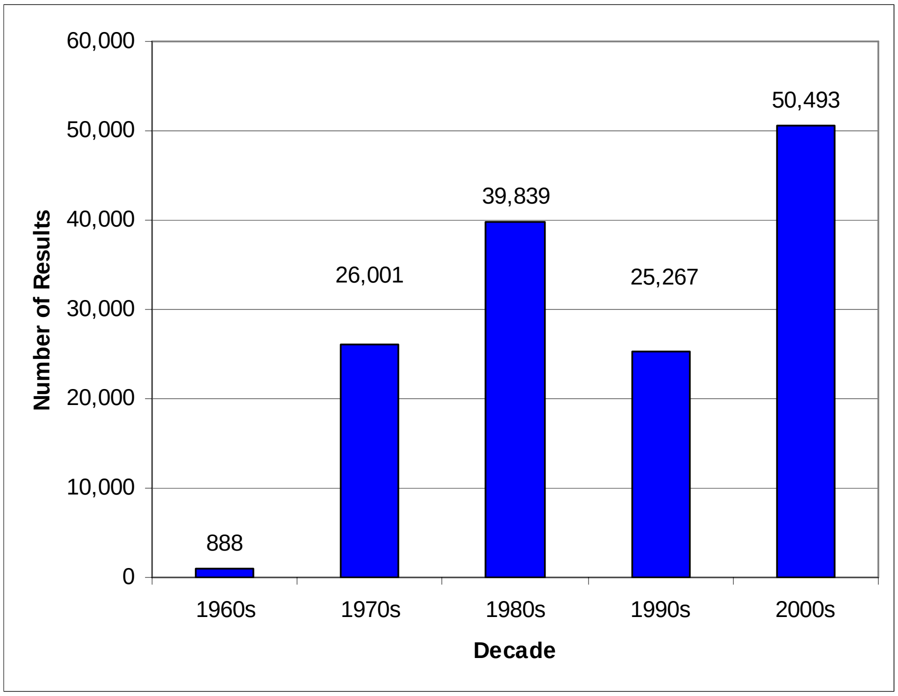


Figure 4.26 SWQM Results by Decade: Nueces River Basin, #21

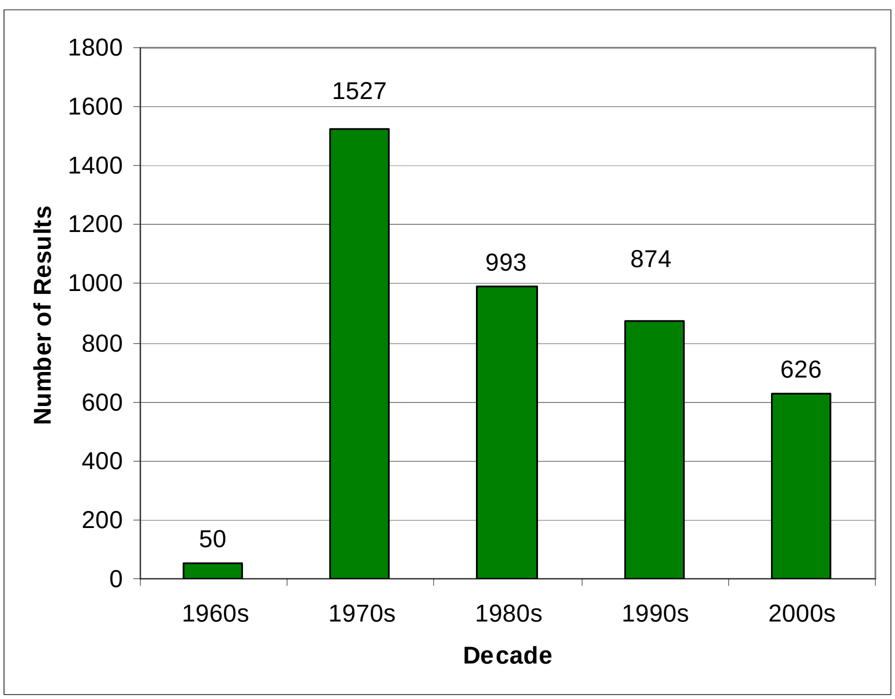


Figure 4.27 SWQM Results by Decade: Lavaca-Guadalupe Coastal Basin, #17

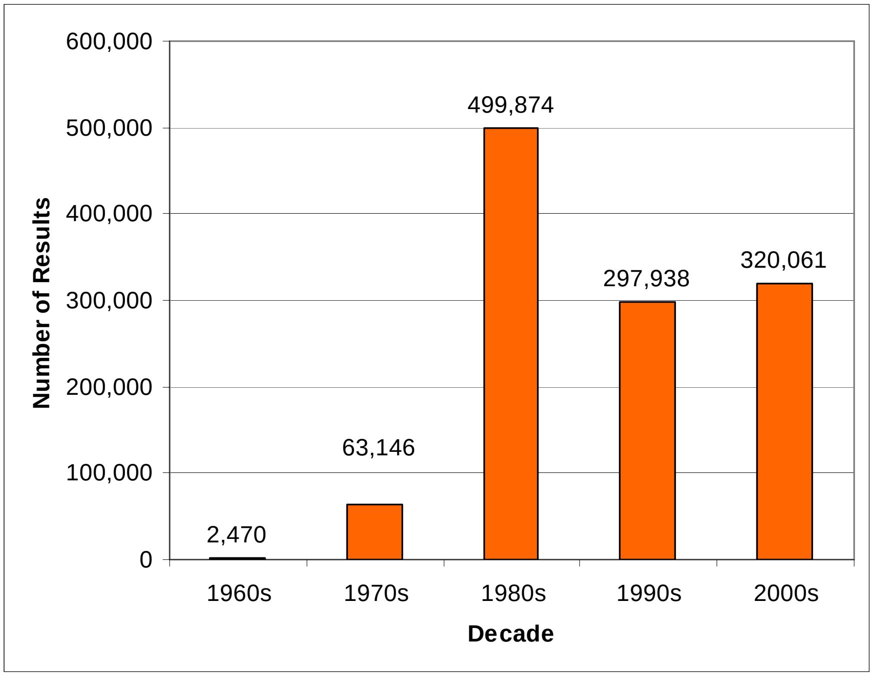


Figure 4.28 SWQM Results by Basin: Colorado River Basin, #14

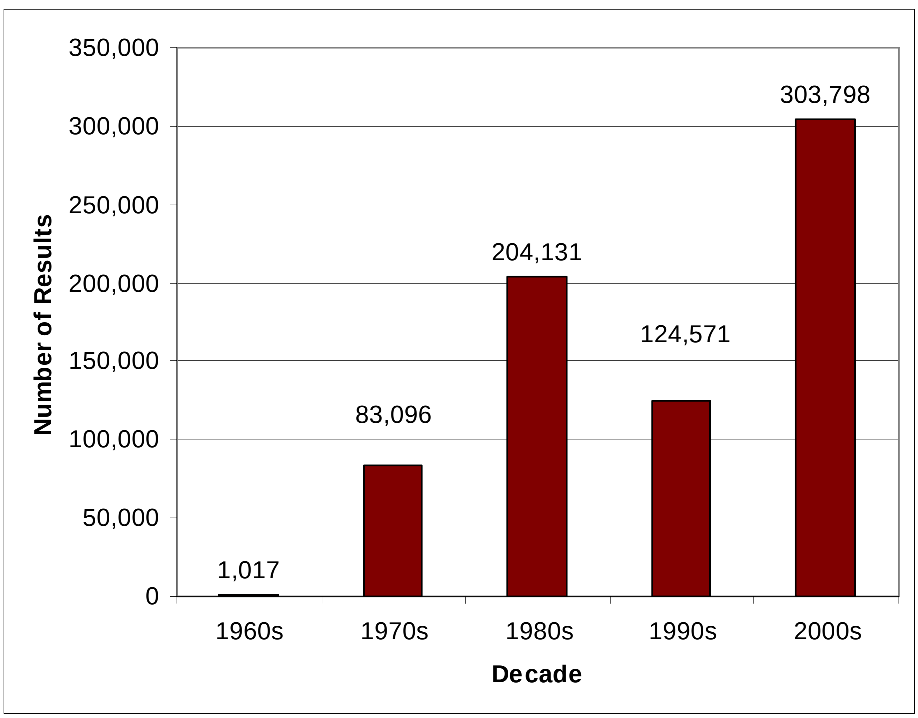


Figure 4.29 SWQM Results by Decade: San Jacinto River Basin, #10

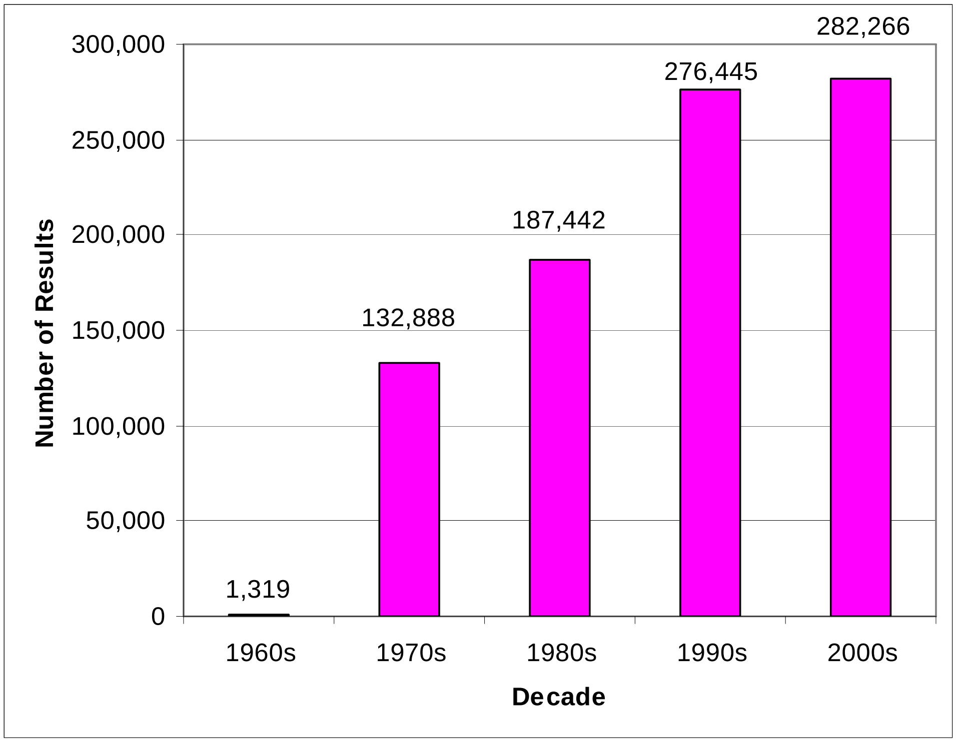


Figure 4.30 SWQM Results by Decade: Trinity River Basin, #8

4.6. Recommendations for Future Work

The current Texas HIS prototype is just that, a prototype. The following is a summary of recommendations for future work developing the Texas HIS. It is essential that future development be coordinated in parallel with the national HIS. However, individual customization, such as the addition of spatial data and web services, is also essential.

4.6.1. Reference Data to NHD

A strength of the HIS is the ability to locate data within a geospatial context. One of the primary features of this spatial context is the NHD, be it NHDPlus or NHDH. The NHD provides a national system for addressing points along water bodies to the water bodies themselves. In order to complete the hydrologic picture, point time series data needs to be referenced to the landscape. While this is already done by using Latitude and Longitude, it makes sense to also do this using an NHD addressing system. Thus, it is recommended that all hydrologic data collection stations be referenced to the NHD. Both data that already exists within the TexasHIS prototype and data that will be added should be referenced to the NHD.

Stations that occur along water bodies such as water quality sampling stations and flow gages should be given an address corresponding to the unique identifier of the waterbody, ReachCode, in the NHD along which the station is located. Completing such an addressing system for the SWQM data already used in the Texas HIS would increase its utility significantly. A similar task of addressing USGS gages was included as part of the creation of NHDPlus. Likewise, the USGS is also addressing dams to their corresponding water bodies.

Stations that do not occur along water bodies such as precipitation and climatic stations should be referenced to the NHDPlus Catchment or NHD Subbasin (also called the 8 digit HUC) in which those stations occur.

4.6.2. High Resolution NHDPlus

While the current version of NHDPlus (scaled at 1:100,000) is a great tool for hydrologic science, a more resolute version is recommended. An NHDPlus High Resolution (1:24,000) has already been discussed for some regions of the United States and of Texas. Once such a product exists, it should be added to the Texas HIS and be the primary hydrographic means from which data is referenced.

4.6.3. Addition of Data Sources

One of the primary improvements recommended for the Texas HIS is the addition of data sources. The utility of a statewide HIS is dependent on the breadth of data that is made available, and on the ease of which that data is available. The current prototype that has been presented includes a limited amount of hydrologic information. However, the framework for adding data has been established. The flexibility of the ODM as a standard data format means that additional sources can be incorporated into the statewide HIS. Also, data can be added through the creation of new web services. While each new data source presents its own unique difficulties with respect to loading it into the ODM or creating a new web service, a few common steps exist. This methodology was employed when moving the SWQM database to the ODM format, described in section 4.4. The lists developed in Table 1.2 through Table 1.5, and especially in Table 1.4 provide a summary of data sources that can be added. Of these, specific attentions should be paid to the WIID, soils data, geologic data, and climatic data.

Before moving data into an ODM database or creating a new web service to an existing database, the source data structure must be thoroughly understood. This requires investigating the current method of data discovery, the current format of data delivery, and an understanding of the current method of data storage. The following questions should be considered: What parameters are necessary to access the data? What type of file is delivered? What fields or tables are returned? How do the fields and tables returned match up with the fields and tables in other observational databases (such as the ODM)? Once these questions are answered, the best approach to the issue of connecting the data source to the HIS server can be addressed. If the current method of data discovery is robust and offers significant flexibility, the creation of a new web service to the existing data source may be most appropriate. If the current method of data discovery is static and inflexible, then the data may need to be moved to a more robust data model. Depending on the original model, the ODM may be the best choice for this, especially since a web service for the ODM already exists.

4.6.4. Addition of HIS Networks

As the HIS becomes more popular as a method of hydrologic information sharing, the development of additional stand-alone HIS networks within Texas is inevitable. At the time this document was written, an HIS test bed was being developed by the scientific research community in Corpus Christi Bay. As the Corpus Christi Bay HIS is further developed, it can be linked to the Texas HIS much like the Texas HIS can be linked to the national HIS. In this way, a complete network of hydrologic information can be developed, much like that shown in Figure 1.6.

Chapter 5. Conclusions

From the statistic in Chapter 1 that 36 percent of hydrologic data users polled spend more than 25 percent of their time collecting data, it is clear that a new method of hydrologic data management is necessary. This is especially true in a state like Texas where large amounts of hydrologic information are produced by state agencies as well as by the numerous academic research institutions. The development of a statewide Hydrologic Information System is proposed as the best solution to hydrologic data management. The strengths of the statewide HIS include integration with systems of varying scales such as the national HIS and a local observatory HIS, the flexibility to add many different types of data sources into various thematic layers, and the ability to discover hydrologic data within a geospatial context. Integration with other HIS's not only makes new data sources available to the state user, but also creates strength in numbers in terms of application development. Applications built on the national HIS can be used with a statewide HIS with little to no modification, given that both systems adhere to the same standards.

The integration of hydrology and information science into the field of hydroinformatics is confusing to both those versed in hydrology, and to those versed in information science. The creation of an HIS is the bridging of these two fields. The understanding of concepts such as an information model, database architecture and web services, as well as concepts such as streamflow, precipitation, and evaporation are all critical. While this document does not perform an exhaustive review of either topic, it touches a few key points that help tie both fields together.

The use of the Observations Data Model has been proposed as the primary method to store hydrologic time series information. Like the HIS itself, this data model is flexible, and can accommodate numerous types of hydrologic information. Great care has been taken in the creation of this model to ensure not only flexibility, but also preservation of data integrity. Also, like the HIS itself, a great strength of the ODM is its wide use. When tools and applications are built on the ODM by one developer, the same tools and applications can be used on all ODM databases with little to no modification given the same database structure.

The creation of a Texas HIS prototype is an important step both for the state of Texas and for the national HIS community. The mission and capabilities of the Texas Natural Resources Information System make it a natural candidate for such a prototype. This prototype will make previously unavailable or unusable hydrologic information available in an easy to use interface that all can access. Additionally, it will pave the way for other statewide and regional systems of a similar nature to be created and integrated, creating a more complete hydrologic picture. The inclusion of Surface Water Quality Monitoring data within the spatial context of the National Hydrography Dataset provides a glimpse at what is possible with a statewide HIS. The framework set forth in this document provides a guide on how to add data sources, and how to further develop this system into an essential tool with extensive utility.

5.1. RECOMMENDATIONS

A key to the success of the HIS concept and to the Texas HIS in particular is the extension of the work set forth in this document, and the extension of the prototype Texas HIS. While recommendations for future work regarding HIS in general are discussed in depth in section 3.12 and for the Texas HIS specifically in section 4.6, those recommendations are summarized here.

The utility of an HIS is dependent in a large part on the data that is made available by this system. Thus, a key to its success is addition of more and more data to the HIS.

The same concept also applies to the ODM. Thus, it is recommended that both the HIS and ODM be more widely adopted throughout the hydrology community.

Another factor leading to the success of the HIS is the HIS server and associated tools. Multiple recommendations have been made for improvements to HIS server, including the ability to serve spatial as well as time series data. The availability of spatial data should not be limited to two dimensions, but should also include three and four dimensional data, such as that in the NetCDF format. Additionally, tools that allow the user to better understand, locate and use the data within the HIS should continue to be developed. One example of such a tool is the NWIS Analyst graphing tool developed at Utah State University. Additionally, standard queries that can be run on data with the ODM would also be extremely useful.

While this document sets forth a prototype and framework for the Texas HIS, there is still much that can be done to make this a successful reality. One recommendation for the Texas HIS is referencing existing and future data sampling sites to the NHD. Doing this will create a direct link between time series values and a hydrologic geospatial context.

Another recommendation for the Texas HIS is the inclusion and integration of additional data. It is recommended that datasets such as the WIID, and those representing themes such as soils, geology, and climatology be added. The addition of new datasets is not trivial. However, the example of adding the SWQM dataset from the TCEQ to the HIS by migrating its contents to the ODM (see section 4.4) shows how this process can work. Another method of adding data to the Texas HIS is linking the Texas HIS to other HIS's from local jurisdictions and individual observatories. This too is recommended as a method of spatially integrating hydrologic data across Texas.

Adding data from other organizations and agencies has the potential to be slowed by limitations of bureaucratic inertia. It is recommended that communication between agencies be made a priority in order to increase and maintain data availability.

Addressing these recommendations for future work will go a long way advancing hydrologic science by thematically integrating hydrologic data in a meaningful, easy to use manner, and presenting this data in a spatial context.

Appendix A: HIS Server Documentation

The following is the most current HIS Server documentation available at the time this thesis was written. It should be noted that this document is still in draft form. Thus, the CUAHSI HIS team (http://www.cuahsi.org/his/) should be consulted for the most recent addition when HIS Server is actually implemented. This document is included here to provide a reference for those interested in actually implementing and installing HIS Server.

Hydrologic Information System Installation and Customization Guide

Draft

February 2007

Prepared for:

CUASHI

Prepared by:

ESRI
380 New York Street
Redlands, California 92373-8100

1.0 Prerequisites

- Internet Information Services (IIS)
- Microsoft .NET framework 2.0
- Visual Studio 2005
- ArcGIS Desktop 9.2
- ArcGIS Server 9.2 .NET Edition (.NET ADF)
- ArcSDE for SQL Express 9.2
- WinZip

2.0 Acquire Microsoft Products

****Note: If you are purchasing a new server for the HIS appliance, you may save a lot of frustration and costs by having the operating system pre-installed by the computer manufacturer. Please see the steps below for more details on which operating system is recommended for the appliance.**

Once you have the proper hardware in place, the very first step will be to install the appropriate operating system and other Microsoft products. For academic institutions, the easiest and most economical method to acquire this software is to negotiate an agreement with Microsoft via the MSDN Academic Alliance Program (MSDN-AA). The process requires a few documents to be sent to Microsoft that basically prove you are an academic institution. Once you are approved and you pay your yearly subscription, you will be sent a set of binders containing Microsoft product CDs or DVDs. Also, you will have the ability to download most of Microsoft's products via the website at http://msdn.microsoft.com/academic following the links for the administrator's section of Microsoft Subscriber Downloads.

The MSDN-AA registration and approval process can be started at https://registermsdn.one.microsoft.com/msdnaa/aa/newstep1.aspx. However, first you may want to look at the overview of MSDN-AA at http://msdn.microsoft.com/academic/program/overview/ and you may also want to look at the usage guidelines shown at http://msdn.microsoft.com/academic/program/usageguide/ to make sure you qualify before proceeding.

Once you have your MSDN-AA software, install the following products in the order listed below:

1) Windows Server 2003 R2 Standard x64 Edition

2) Internet Information Server (IIS) – This is a component of Windows Server 2003, however, if it was not loaded during the initial install, it can be added by going to Start→Settings→Control Panel→Add or Remove Programs→Add/Remove Windows Component

3) Visual Studio .NET 2005 Professional

3.0 Install ArcGIS Products

Install the ArcGIS products in the following order (highly recommended):

1) ArcGIS 9.2 Desktop

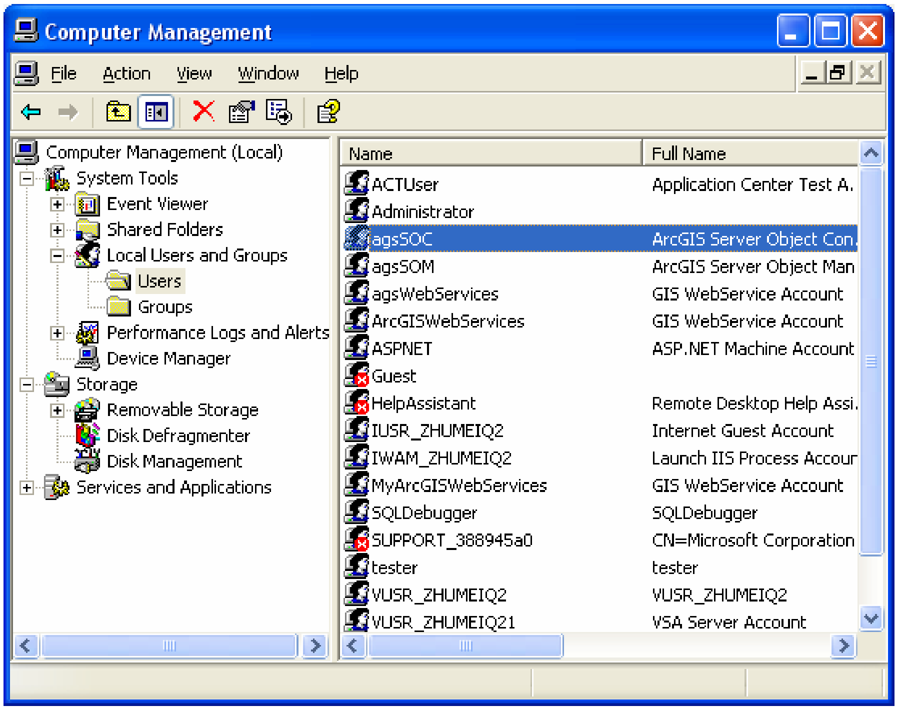
2) ArcGIS Server 9.2 DotNet Edition

3) ArcSDE/SQLExpress Personal version

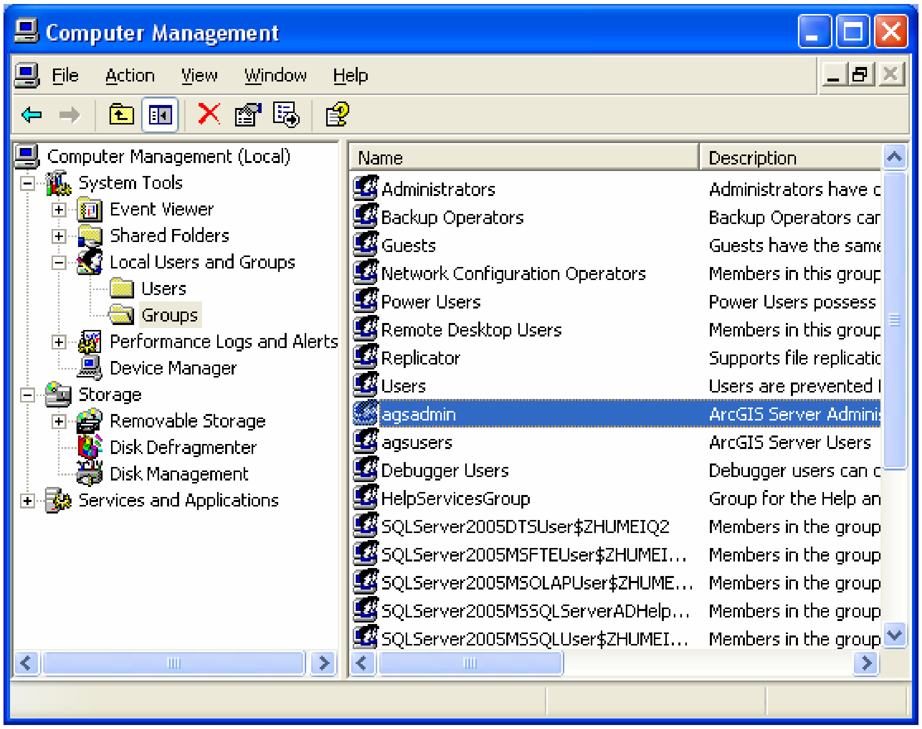
4.0 Verify ArcGIS User Accounts

After installing ArcGIS Server 9.2, if you correctly follow the directions for post installation, you should have created the user accounts which are necessary to run the server, and granted them the necessary privilege on the system. By default, two user accounts are created which are agsSOC (or arcgisSOC) and agsSOM (or arcgisSOM). To verify these two accounts have been created and appropriately configured, follow the following steps.

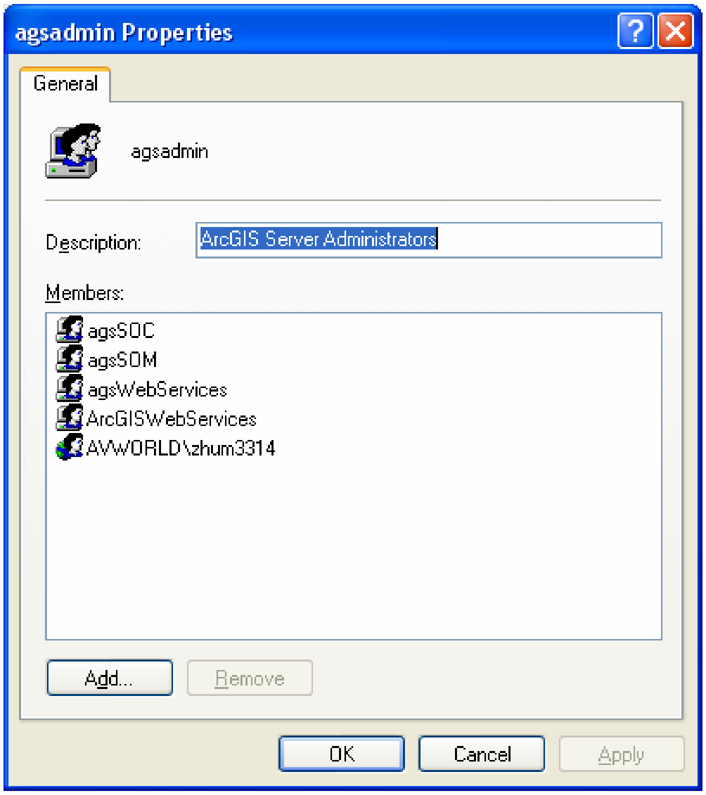
1. Open you Computer Management from Control Panel and expand Users under System Tools>Local Users and Groups>Users, and you should see agsSOC and agsSOM are listed as below:



2. Open you Computer Management from Control Panel and expand Users under System Tools>Local Users and Groups>Groups, and you should see agsadmin and agsusers are listed as below:



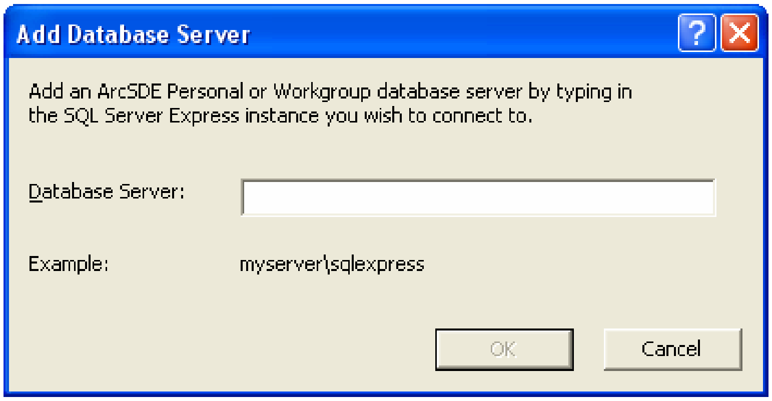
3. Highlight on agsadmin, right click and open the property window. Verify that agsSOC and agsSOM are listed as members in the agsadmin user group.



5.0 Create Database Server Object and Attach Database File

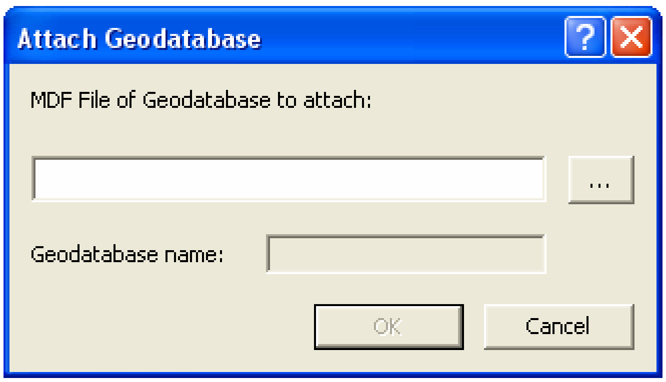
Before creating the database sever object, it is assumed that your ArcSDE for SQL Express 9.2 has been properly installed and configured. In addition, the database file (a .mdf file) has been copied to your local disk.

1. Open ArcCatalog and DoubleClick on Database Servers\Add Database server to create a new Database server. Type in your server name followed by the SQL Express instance name (e.g. seademon\sqlexpress) that you wish to connect to. You should see the server added into Database Servers now. If you have any problems adding the GIS server, make sure your login is part of the agsadmin user group.



2. Double click the newly created database server object to connect to it.

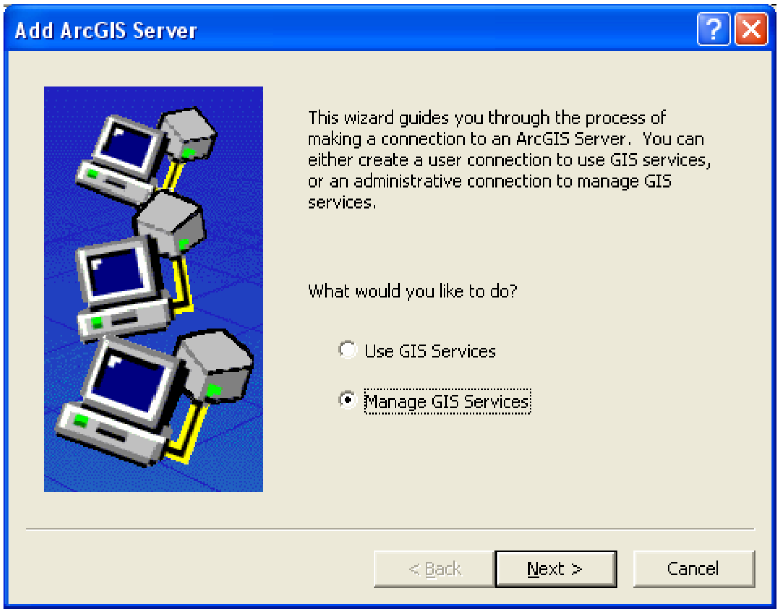
3. Right click on the database server object and select “Attach” to attach the database file to ArcSDE. Click ok when done.



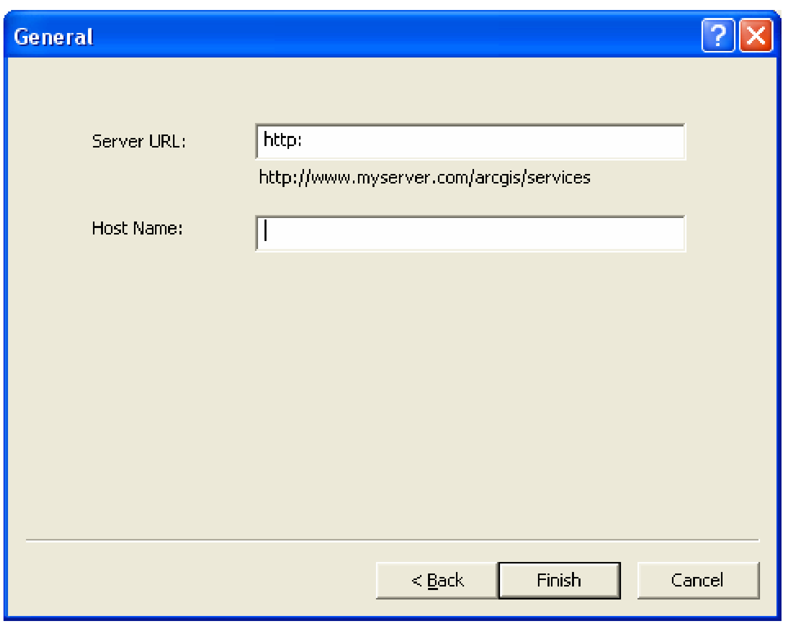
4. Open the ArcMap document provided. Reset the data source for layers. Save and close the document.

6.0 Add ArcGIS Server Object and Create Map Service

1. From ArcCatalog, expand GIS Servers, and click Add ArcGIS Server. In the popup window, select Manage GIS Services and click Next.

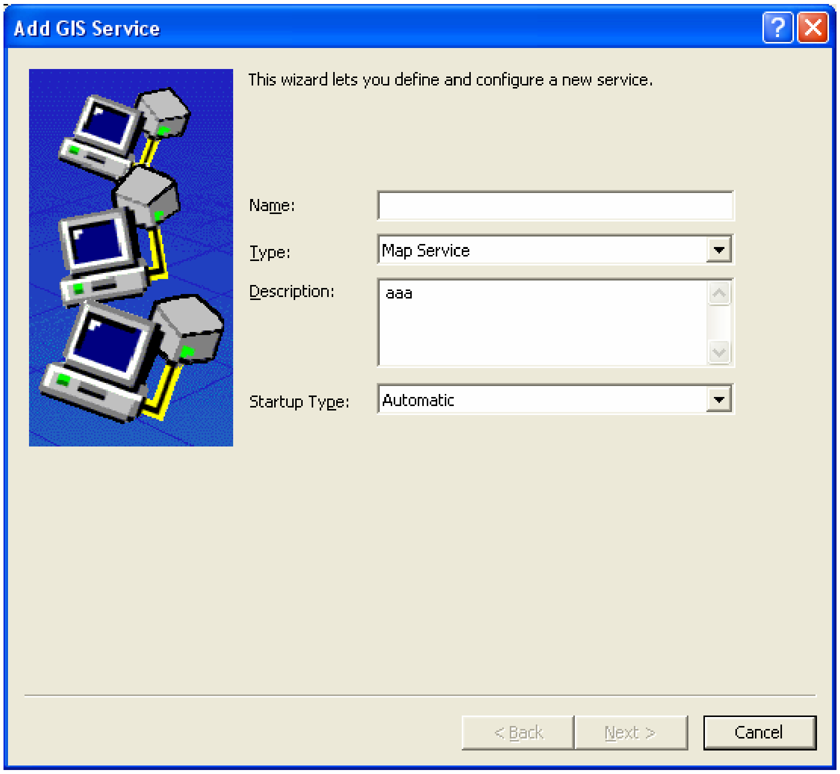


2. Type in the map service server name and click finish. You should have the server object now. Note: if you have problem adding the service, refer to section 4 to verify the setting of SOC and SOM accounts. You may have to rerun ArcGIS Server post installation if problem persists.

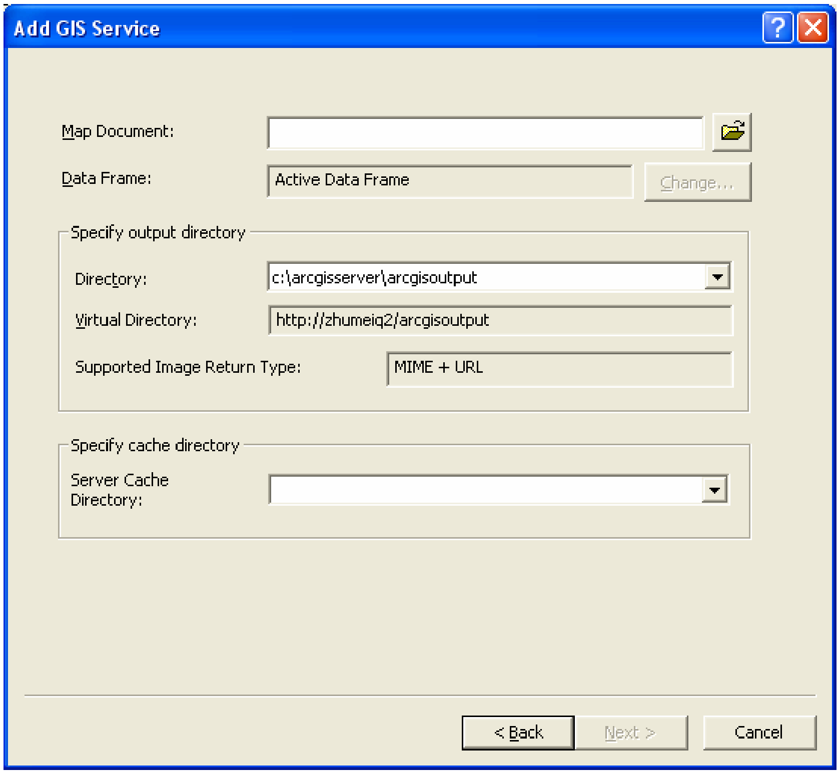


3. Right click on the server object and select "Add New Service". In the popup window, give a name and description of your map service, leave the type of service as default ("map service") and click next. Note: you will use the name of the service when

configuring the Resource Manager of the web application, which is described in section 7.



4. Browse to the ArcMap document you wish to publish over the web, and choose c:\arcgisserver\arcgiscache as the server cache directory and click next.



5. Leave everything as default in the following windows and finish the creation of your map service. You should have the service listed under your server object now.

7.0 Copy and Edit the .NET Web Application File

1. Go to C:\Inetpub\wwwroot and create a folder named HIS.

2. Copy the HIS program source code and associated files and file folders to C:\Inetpub\wwwroot\HIS.

3. Open Computer Management from Control Panel, and navigate to Internet Information Services\Web Sites\Default Web Site\HIS.

4. Right click on the folder of HIS and select property.

5. In the property window, click ‘Create’ to make this folder a web site.

6. Go to C:\Inetpub\wwwroot\HIS and make the entire folder writable.

7. Open Microsoft Visual Studio 2005 and click File>Open Web Site to pen HIS folder.

8. Double click on web.config to open the file and find the <Identity> node. Delete the node. This node looks like the following.

```
<identity
  configProtectionProvider="DataProtectionConfigurationProvider">
 <EncryptedData>
  <CipherData>

    <CipherValue>AQAAANCMnd8BFdERjHoAwE/Cl+sBAAAA4GMXGBsNBkCAuiqqBdp
    anAQAAAACAAAAAAADZgAAqAAAABAAAAB7rLZl0rpxoeyMm3op8OLUAAAAAASAAAC
    gAAAAEAAAAGqlJzUgKhSYzrSrKMre8gKgAAAAQ0ZgLvxuyNnIpbpKY5ywlImYo8E
    ccF+r4QIZDAqU3o5e2BRCXphic31ABY6TwQOyz14kzsvR1xaFmLz13iW/OW0gLAn
    nMDQgsOBYbE6Ocn/02EtSzPVF5jivfSA4NaN5nVOcKouVu9yU1avE/X9y9eV9w5E
    z8aGbT6+nTill9AQrBPrpj1trxgqBe/nk1cyO9vBkAuOQ8itrQUvSQ12H2RQAAAC
    V6G5y3HeyXY7HUVVm89Obo1cWMw==</CipherValue>
  </CipherData>
 </EncryptedData>
</identity>
```

9. Make sure you close this file by clicking the X on the upper right.

10. Right click at the project level (the very top in your solution explorer window) and click “Add ArcGIS Identity”.
    1. Use the windows logon which has admin right to fill the box in this window.
    2. Click ok.

11. Double click on default.aspx to open it.

    1. Find the MapResourceManager and click on the tiny arrow on the upper right side of it to open a MapResourceManager Tasks window.

    2. Click on Edit Resources to open the Collection Editor window.

    3. Find Definition in the Information section and click on the little “…” button at the right end of the definition text. This will open the Definition editor.

    4. In Definition editor window, Data Source is the server name, Resource is the name of the map service created in step 6.

    5. Click OK.

12. Save your changes. Now you can run the program.

8.0 Edit the XML Configuration File and Customize HIS Web Application

The HIS web application is designed as a generic application that can be easily configured to access hydrological data from diverse data sources through a map viewer. A published ArcMap document that serves as the map service for the HIS application, and an XML based configuration file together define what layers to be displayed, what data provider to be included, and what attribute of a feature layer to be queried and displayed in the map viewer. Both the ArcMap document and XML file can be edited by the user. When starting the HIS web application, the XML configuration file will be first read and interpreted to establish data access channels between the map viewer and individual data providers. The design of such an XML configuration file makes the application flexible at accessing different data sources as needed by different application hosts.

In the ArcMap document, each data provider is represented as an individual feature layer. For example, NWIS_DailyValue feature layer represents the NWIS site locations that measure parameters of daily values. The time series of an observation can be obtained by passing to the NWIS web service the site identifier, which is an attribute of the NWIS_DailyValue feature layer. Such information also needs to be reflected in the XML file. On the other side, information defined in the XML configuration file must be consistent with the ArcMap document. For a particular data provider, the network name, web service URL, unique identifier field, and map viewer displaying fields can be fully defined in this XML file.

8.1 Structure of the XML Configuration File

The basic structure of the XML file is:

```
<HISNetworks>
   <HISNetwork>
      <ApWebFields>
           <ApWebField>
                              … detail definition of ApWebField …
          </ApWebField>
                    … More ApWebField node can follow to define details for additional displaying fields …
      <ApWebFields>
   </ HISNetwork>
   … More HISNetwork node can follow to define details for additional networks …
</HISNetworks>
```

<HISNetworks> node is the root node of the XML file and acts as a collection of all the <HISNetwork> nodes, with each to support an individual data provider. For example, a <HISNetwork> node is needed to support data access from the NWIS Daily Value web service, and a separate <HISNetwork> node is required to support data access from EPA STORET web service.

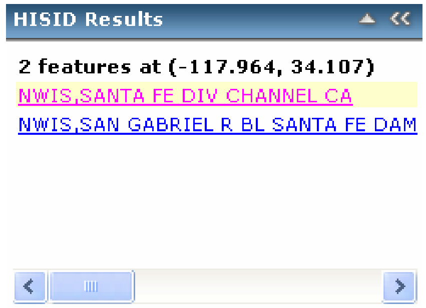
A few attribute nodes are designed for <HISNetwork> node, which provide detail information to the web application for establishing the communication between the map viewer and individual data providers.

- *Script:* name of the JavaScript function to call.
- *RecordSetXML:* not used now. Reserved for future use.
- *Name:* name of the network, such as NWIS, EPA, etc.
- *TagName:* tag name of the network, which is used exclusively by the application. Usually it's the name of the network.
- *LayerName:* name of the network feature layer as shown in the ArcMap document.
- *LayerIndex:* not used now. Reserved for future use.
- *SiteCodeField:* field name in the feature layer that is used as unique identifier of a feature. Value of this field is passed over the web service to obtain data for a feature.
- *WebServiceURL:* URL of the web service.

Among these attributes, *Script*, *Name*, *LayerName*, *SiteCodeField*, and *WebServiceURL* are required for each <HISNetwork> node. The rest are optional. Below is a sample <HISNetwork> node with attributes. This node tells the HIS web application to access NWIS data by using the web service at "http://water.sdsc.edu/WaterOneFlowDev/NWIS/DailyValues.asmx", and by using the feature layer named "nwis_DailyValue" in the published map document. "SiteCode" is the field name in feature layer "nwis_DailyValue" that stores unique identifier of features. Value of this field will be retrieved and passed to the web service to get data for a feature. Also, "GetSiteInfo" is the JavaScript function used to perform such a call to the web service.

```
<HISNetwork Script="GetSiteInfo" RecordSetXML="" TagName="NWIS" Name="NWIS"
LayerName="nwis_DailyValue" LayerIndex="0"
WebServiceURL="http://water.sdsc.edu/WaterOneFlowDev/NWIS/DailyValues.asmx"
SiteCodeField="SiteCode">
</ HISNetwork>
```

Within each <HISNetwork> node, there are usually 2 <ApWebFields> nodes. One is designed for providing field names that are used as parameters passing to the called function or web services, and the other is for attributes to be retrieved from the feature layer and displayed on the map viewer, as shown in the picture below. Such attributes are concatenated according to the order specified in <ApWebFields> node and displayed as hyperlinks. When clicked, underlying parameters are to be passed to the called function or directly to web services.



A few attribute nodes are designed for <ApWebFields>.

- *Desc:* description of the purpose of the ApWebFields node, which can be either *ParameterFields* or *DisplayFields*. *ParameterFields* provides field names used as parameters passing to the called function, while *DisplayFields* provides field names to be displayed on the map viewer.
- *Name:* name of ApWebFields. which can be either *ParameterFields* or *DisplayFields*.
- *TagName:* tag name of of ApWebFields that is exclusively managed by the program. Usually this is the name of ApWebFields.
- *IsReadOnly:* indicates if the field is read only or writable.
- *IsFixedSize:* indicates if the displaying size for a field is fixed or not.
- *IsSynchronized:* indicates if the field needs to be synchronized.
- *Count:* number of ApWebField nodes within the current ApWebFields node.

Among the above sttributes, *Name* is required by the HIS application. The rest are optional. Below is a sample <ApWebFields> node containing field information used as parameters passing over the web services.

```
<ApWebFields Desc="ParameterFields" Name="ParameterFields" TagName="ParameterFields"
IsReadOnly="False" IsFixedSize="False" IsSynchronized="False" Count="2">
</ApWebFields >
```

A <ApWebFields> node containing displaying field information has different "Name" attribute. Here *ParameterFields* and *DisplayFields* are two keywords designed by the system and they are case sensitive.

```
<ApWebFields Desc="DisplayFields" Name="DisplayFields" TagName="DisplayFields"
IsReadOnly="False" IsFixedSize="False" IsSynchronized="False" Count="2">
</ApWebFields >
```

Each <ApWebFields> node acts as a collection of <ApWebField> nodes that provide more detail information of a field, such as field name and field alias. A <ApWebFields> node can have one or more <ApWebField> nodes. A few attribute are designed for <ApWebField> node.

- *ConstValue:* indicates if a constant value exists for this field, which can be either "True" or "False". If ConstValue equals to "True", then the field value will not be read on the fly. Instead, value of the "Name" attribute will be used.
- *FieldOrder:* the order of the field for displaying or for passing as a parameter.
- *AllowEdit:* indicates if this field is editable.
- *Desc:* description of the field.
- *FieldAlias:* field alias.
- *Name:* field name.
- *TagName:* tag name of the field, which is exclusively managed by the program.
- *Order:* not used now. Reserved for future use.
- *Type:* field type, which is read from the ODM on the fly by the program.
- *Text:* not used now. Reserved for future use.

Among the above attributes, *ConstValue, Name* and *FieldOrder* are required by the HIS application. The rest are optional. Below is a sample <ApWebField> node.

```
<ApWebField ConstValue="False" FieldOrder="1" AllowEdit="1" Desc="SiteName"
FieldAlias="SiteName" Name="SiteName" TagName="SiteName" Order="0" Type="0"
Text="SiteName" />
</ ApWebField >
```

Now we can look at a complete XML file that has all the required structures described above. The following XML demonstrates a basic structure that contains only one <HISNetwork> node which provides necessary information in regard to accessing NWIS network. The NWIS daily value is available at "http://water.sdsc.edu/WaterOneFlowDev/NWIS/DailyValues.asmx", and can be queried by using the feature layer named "nwis_DailyValue" in the published map document. "SiteCode" is the field name in this feature layer that stores unique identifier of features. Value of this field can be retrieved and passed to the web service to get data for a feature. When calling the web service, values of SiteCode and NetworkNam are to be retrieved from the feature layer and are passed as parameters when calling the web service. Meanwhile, values of SiteCode and NetworkNam are concatenated and displayed on the map viewer. In this example, parameter fields and displaying fields are the same, but they can be different in a real application. "GetSiteInfo" is the JavaScript function used to perform such a call to the web service.

```
<HISNetworks IsReadOnly="False" IsFixedSize="False" IsSynchronized="False" Count="1">
   <HISNetwork Script="GetSiteInfo" RecordSetXML="" TagName="NWIS" Name="NWIS"
    LayerName="nwis_DailyValue" LayerIndex="0"
    WebServiceURL="http://water.sdsc.edu/WaterOneFlowDev/NWIS/DailyValues.asmx"
    SiteCodeField="SiteCode">
      <ApWebFields Desc="ParameterFields" Name="ParameterFields"
      TagName="ParameterFields" IsReadOnly="False" IsFixedSize="False"
      IsSynchronized="False" Count="2">
         <ApWebField ConstValue="False" FieldOrder="0" AllowEdit="1"
            Desc="SiteCode" FieldAlias="SiteCode" Name="SiteCode"
            TagName="SiteCode" Order="0" Type="0" Text="SiteCode" />
         <ApWebField ConstValue="False" FieldOrder="1" AllowEdit="1"
            Desc="NetworkNam" FieldAlias="NetworkNam" Name="NetworkNam"
            TagName="NetworkNam" Order="0" Type="0" Text="NetworkNam" />
      </ApWebFields>
      <ApWebFields Desc="DisplayFields" Name="DisplayFields"
      TagName="DisplayFields" IsReadOnly="False" IsFixedSize="False"
      IsSynchronized="False" Count="2">
         <ApWebField ConstValue="False" FieldOrder="1" AllowEdit="1"
            Desc="SiteName" FieldAlias="SiteName" Name="SiteName"
            TagName="SiteName" Order="0" Type="0" Text="SiteName" />
         <ApWebField ConstValue="False" FieldOrder="0" AllowEdit="1"
            Desc="NetworkNam" FieldAlias="NetworkNam" Name="NetworkNam"
            TagName="NetworkNam" Order="0" Type="0" Text="NetworkNam" />
      </ApWebFields>
    </HISNetwork>
</HISNetworks>
```

When adding a new network to the HIS server application, a new <HISNetwork> node needs to be added into the XML file by following the directions described above.

8.2 Add Local Data or New Network to HIS Web Application

This section provides step-by-step direction on how to add a new network to the HIS web application.

1. Make sure you have the feature class in your geodatabase representing the stations where variables/parameter are observed or measured. E.g. feature class EPA_STOREIT represents EPA point stations that measure water quality parameters.

    i. Make sure there is a unique ID field in the feature class that can be used to identify your stations, and it is also the identifier used by the web service to provide data. E.g. WSSiteID is the unique identifier.

    ii. Select fields whose values need to be displayed on the map viewer. E.g. StationNam.

2. Add this feature class to your ArcMap document which serves as map service and save the ArcMap document. E.g. EPA_STOREIT is added into ArcMap

document and the layer name is EPA_STOREIT.

Note: Make sure the layer is added into the appropriate data frame if you have more than one data frames in the map. It is highly recommended that only one data frame in the map. Also, if you add new layer to an existing map service, make sure you stop it first. You can restart the map service after adding the new feature layer.

3. Make sure there is a working web service from which corresponding observation data can be obtained. E.g. **http://water.sdsc.edu/wateroneflow/EPA/cuahsi_1_0.asmx** is a working web service for EPA_STOREIT.

4. Edit the XML configuration file.

    i. Create a new <HISNetwork> node and give feature layer name, network name, web service URL to appropriate attribute nodes. Give "GetSiteInfo" as the value of Script. Normally you don't need to change the value for this attribute unless you have a customized function especially written for calling this web service in the JavaScript file.

```
<HISNetwork Script="GetSiteInfo" RecordSetXML="" TagName="EPA"
Name="EPA" LayerName="EPA_STOREIT" LayerIndex="1"
WebServiceURL="http://water.sdsc.edu/wateroneflow/EPA/cuahsi_1_0.asmx"
SiteCodeField="WSSiteID">
</ HISNetwork>
```

    ii. Create a new <ApWebFields> node as a child node of <HISNetwork> and give name, description, and number of fields to appropriate attribute nodes. This node is used for passing parameters. Therefore, ParameterFields is given as the value for attributes Desc and Name.

```
<ApWebFields Desc="ParameterFields" Name="ParameterFields"
TagName="ParameterFields" IsReadOnly="False" IsFixedSize="False"
IsSynchronized="False" Count="2">
</ApWebFields >
```

    iii. Create 2 <ApWebField> nodes as child nodes of <ApWebFields> and assign appropriate values to the attributes.

```
<ApWebField ConstValue="False" FieldOrder="0" AllowEdit="1"
      Desc="WSSiteID" FieldAlias="WSSiteID" Name="WSSiteID"
      TagName="WSSiteID" Order="0" Type="0" Text="WSSiteID" />
 <ApWebField ConstValue="True" FieldOrder="1" AllowEdit="1" Desc="EPA"
      FieldAlias="EPA" Name="EPA" TagName="EPA" Order="0" Type="0"
      Text="EPA" Value="EPA" />
```

Usually a network name and a unique ID field are required by the HIS Server application to obtain data over the web. Here, the first <ApWebField> indicates that field WSSiteID is one parameter. However, the second <ApWebField> has a ConstValue equals to “True”. This means rather than reading a field value from the feature class, a constant value, which is “EPA” in this case, is used as the second parameter, which indicates the network name.

iv. Copy the entire <ApWebFields> node and insert it as a new child node of <HISNetwork>. This node provides field names for displaying on the map viewer. Therefore, change ParameterFields to DisplayFields as the value of corresponding attributes. Displaying fields and parameter field can be either the same or different. To keep it simple, we use same fields in this example. Now your new <HISNetwork> node should look like the following.

```
<HISNetwork Script="GetSiteInfo" RecordSetXML="" TagName="EPA"
        Name="EPA" LayerName="epa_STOREIT" LayerIndex="1"
        WebServiceURL="http://water.sdsc.edu/wateroneflow/EPA/cuahsi_1_0.asm
        x" SiteCodeField="WsSiteID">
        <ApWebFields Desc="ParameterFields" Name="ParameterFields"
            TagName="ParameterFields" IsReadOnly="False" IsFixedSize="False"
            IsSynchronized="False" Count="2">
                <ApWebField ConstValue="False" FieldOrder="0" AllowEdit="1"
                    Desc="WSSiteID" FieldAlias="WSSiteID" Name="WSSiteID"
                    TagName="WSSiteID" Order="0" Type="0" Text="WSSiteID" />
                <ApWebField ConstValue="True" FieldOrder="1" AllowEdit="1"
                    Desc="EPA" FieldAlias="EPA" Name="EPA" TagName="EPA"
                    Order="0" Type="0" Text="EPA" Value="EPA" />
        </ApWebFields>
        <ApWebFields Desc="DisplayFields" Name="DisplayFields"
            TagName="DisplayFields" IsReadOnly="False" IsFixedSize="False"
            IsSynchronized="False" Count="2">
                <ApWebField ConstValue="False" FieldOrder="1" AllowEdit="1"
                    Desc="StationNam" FieldAlias="StationNam"
                    Name="StationNam" TagName="StationNam" Order="0"
                    Type="0" Text="StationNam" />
                <ApWebField ConstValue="True" FieldOrder="0" AllowEdit="1"
                    Desc="EPA" FieldAlias="EPA" Name="EPA" TagName="EPA"
                    Order="0" Type="0" Text="EPA" Value="EPA" />
      </ApWebFields>
</HISNetwork>
```

v. Save the XML file and close it. Now you can launch the web application and request data through the new network.

Appendix B: ODM Web Service Documentation

The following document is supplied as part of the Generic OD Web Service provided by CUAHSI as part of the National HIS project, and can be found at http://water.sdsc.edu/genericODws.zip. It was created on February 16, 20007, and is still in draft form. As the ODM and ODM web services continue to evolve, so too may the instructions contained in this document. It is provided here as an introduction and a guide. If the web service here is actually implemented, the user should consult with the CUAHSI HIS team at http://www.cuahsi.org/his/ to check for additional releases.

Introduction

The CUAHSI Observations Data Model (ODM) is a schema for storing hydrologic observations time series data. Scientists apply the schema to create an Observations Database (OD) for their data. To assist scientists in publishing their OD data, CUAHSI has developed an OD web service. This document explains how to install and configure an OD web service for your OD.

Requirements

Hardware:

- HIS Server Appliance

Software:

- Micosoft Windows XP or 2003
- IIS with ASP.Net
- .Net 2.0 Framework
- SqlServer Express, or SQL Server Database that service can connect to

Data:

- Observations Database (.mdf file)
- GenericODws.zip file containing Generic OD web service support files (from CUAHSI). GenericODws.zip contains a sample OD.mdf database for testing purposes.

Personnel:

- User with administrator account

Note: If you decide to use the sample OD.mdf database included in GenericODws.zip, make sure a file called "OD_Log.ldf" does not exist in the App_Data folder. If it exists, delete it.

Procedure

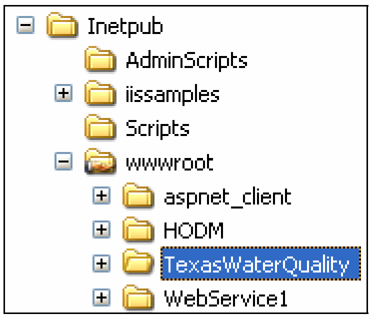
This section describes the steps required to set up the service. For the advanced user, each section begins with a description of what is being done. If you understand the description, you may proceed on your own. For users who need more information than the short description, step-by-step instructions follow the description.

Note: If you have more than one OD, you'll have to set up a separate OD web service for each database.

Unzip the Web Service Folder

First you'll unzip the web service folder to the IIS folder, and give it a name appropriate for your database.

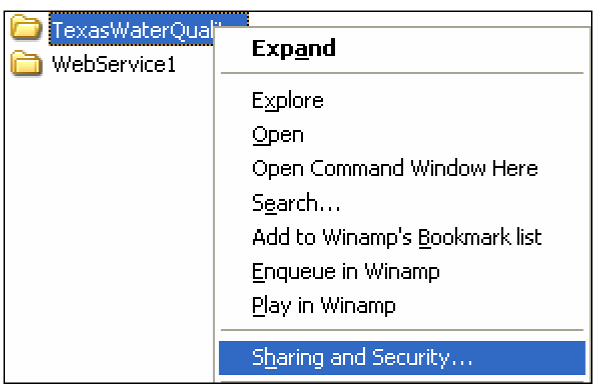
1. **Unzip** GenericODws.zip to the IIS directory (e.g., C:\Inetpub\wwwroot). This places a folder in the IIS directory called "GenericODws".
2. **Rename** the GenericODws folder to the desired name for your service, e.g., "TexasWaterQuality".



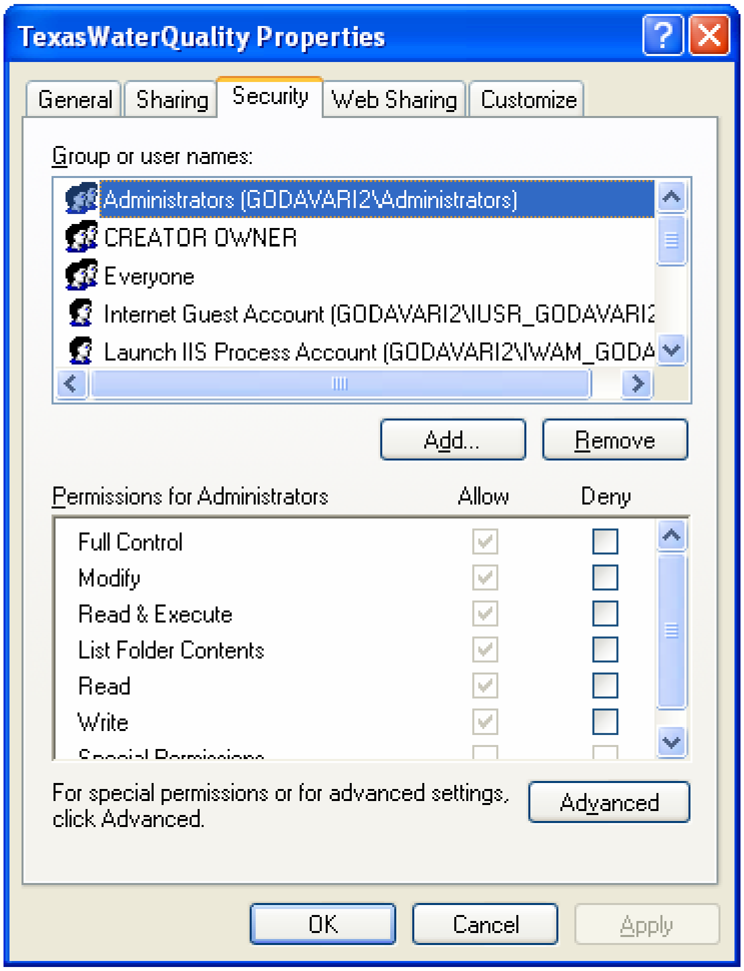
Set Folder Permissions
You must set permissions for the folder so that the appropriate accounts have enough control to make the service work. The permissions you are setting are:

- [Your Service Folder]: Everyone - Read
- [Your Service Folder]: ASPNET - Full Control
- [Your Service Folder]\App_Data: Network_Service - Full Control

1. In Windows Explorer, **right click** on the web service folder, and **click** *Sharing and Security*.



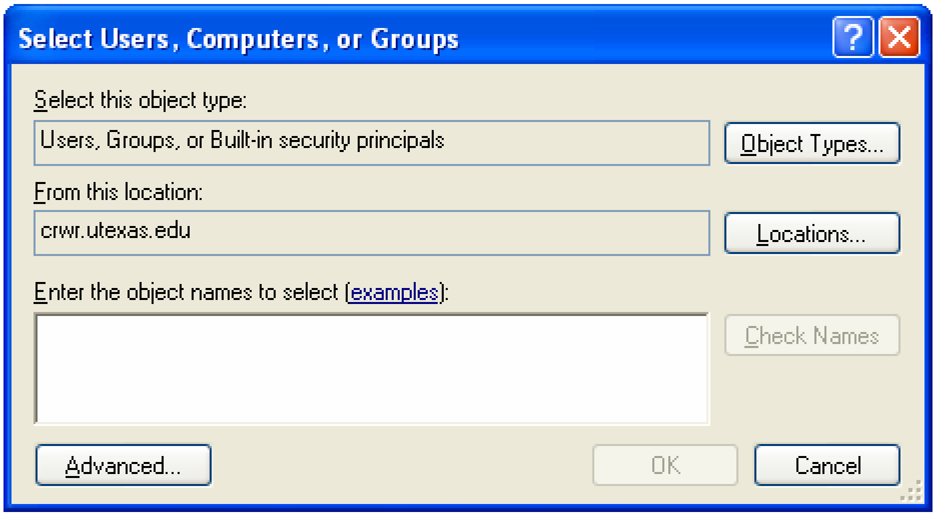
2. **Click** the *Security* tab.



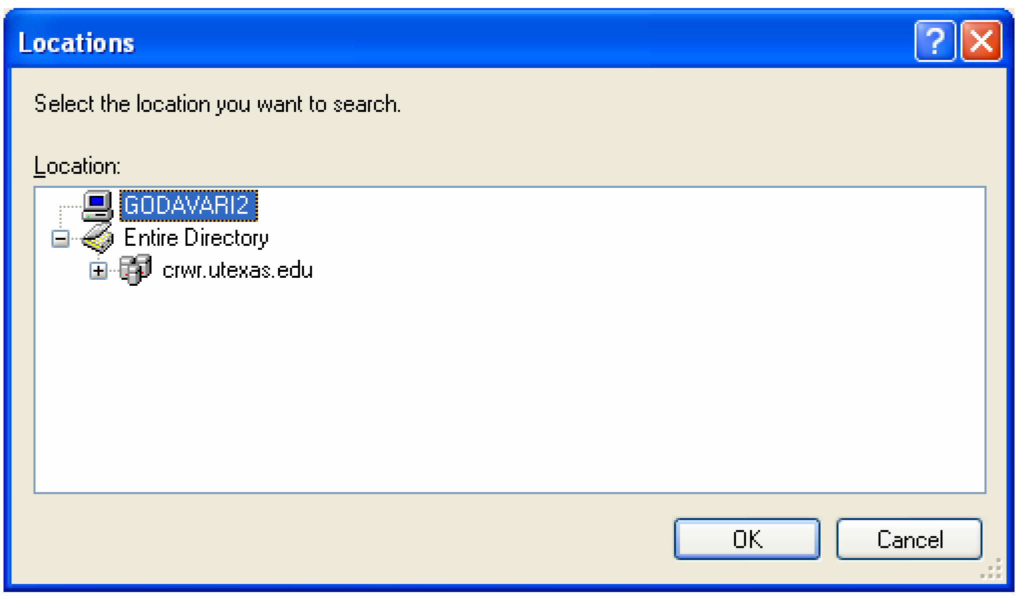
3. Set the following permissions:
    a. Everyone - Read
    b. ASPNET - Full Control

If you did not see the group or user name for the above permissions, you will need to add that permission. Steps 4-11 walk you through that process. Otherwise, skip to step 12.

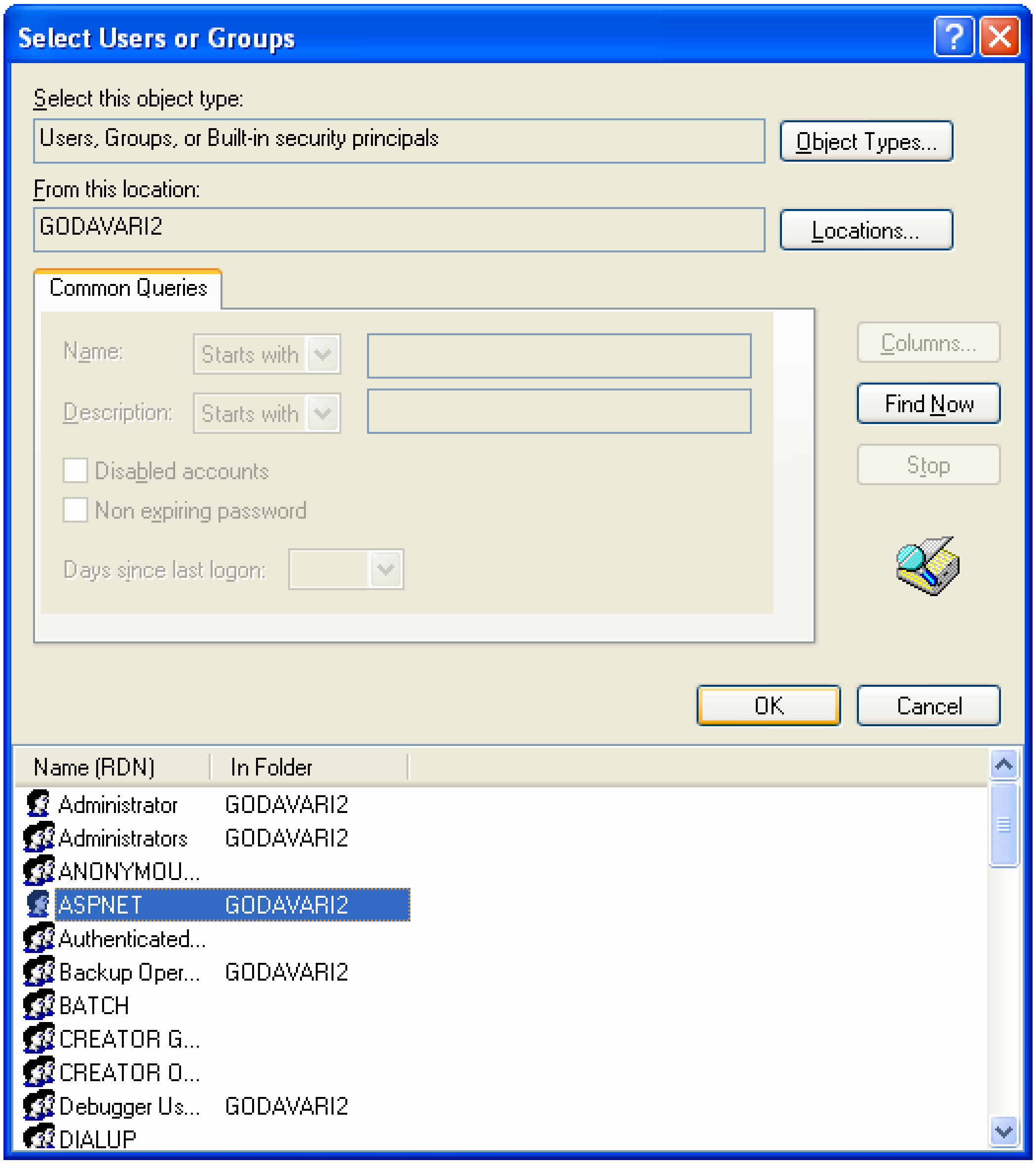
4. **Click** *Add*.
5. In the *Select Users, Computers, or Groups* window, **click** *Advanced*.



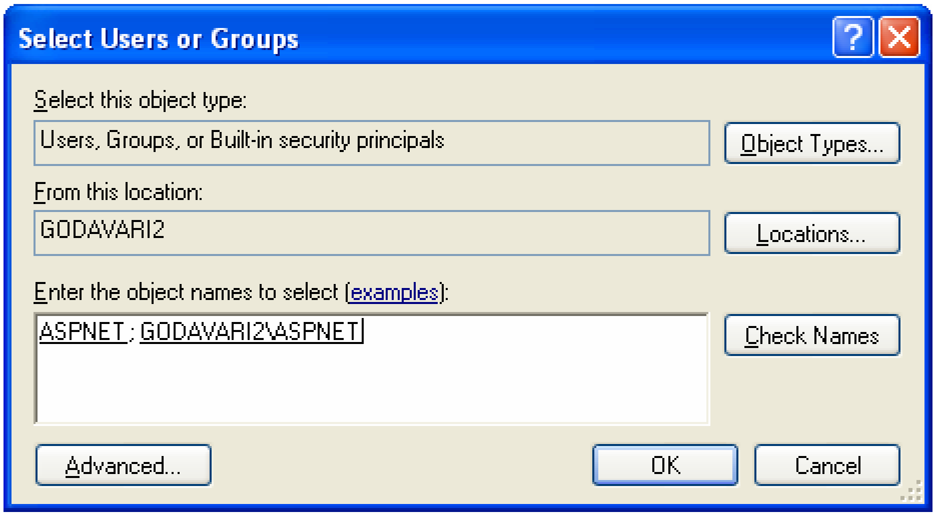
6. In the window that opens, **click** *Locations*.
7. In the window that opens, **select** your computer, and then **click** *OK*.



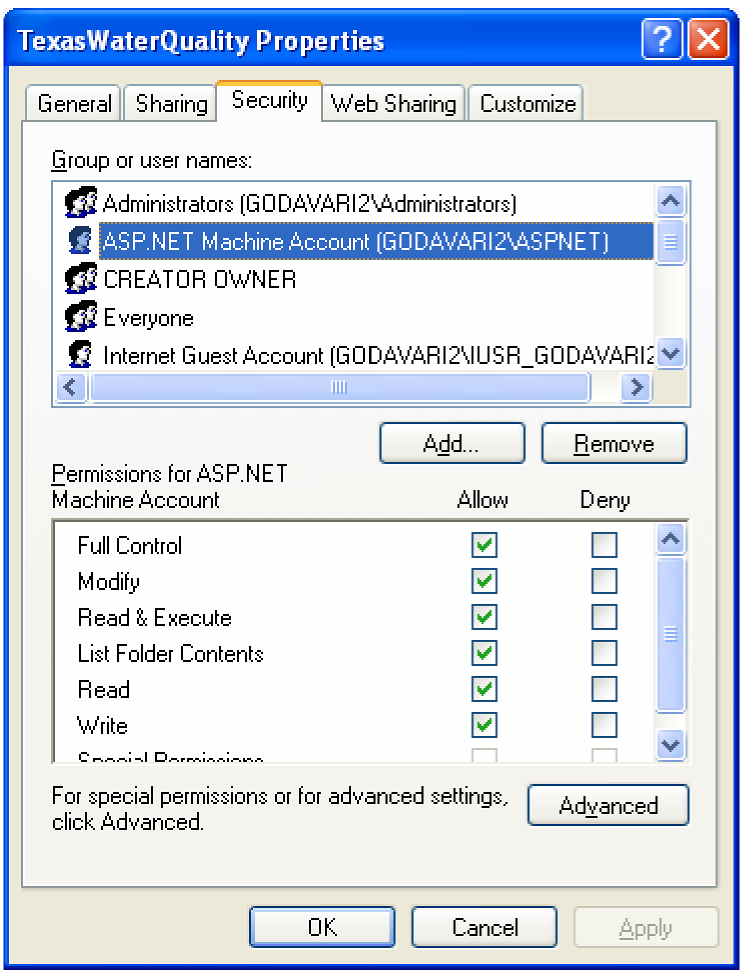
8. **Click** *Find Now*.
9. Scroll to the desired name, highlight it by **left clicking** on it, and **click** *OK*.



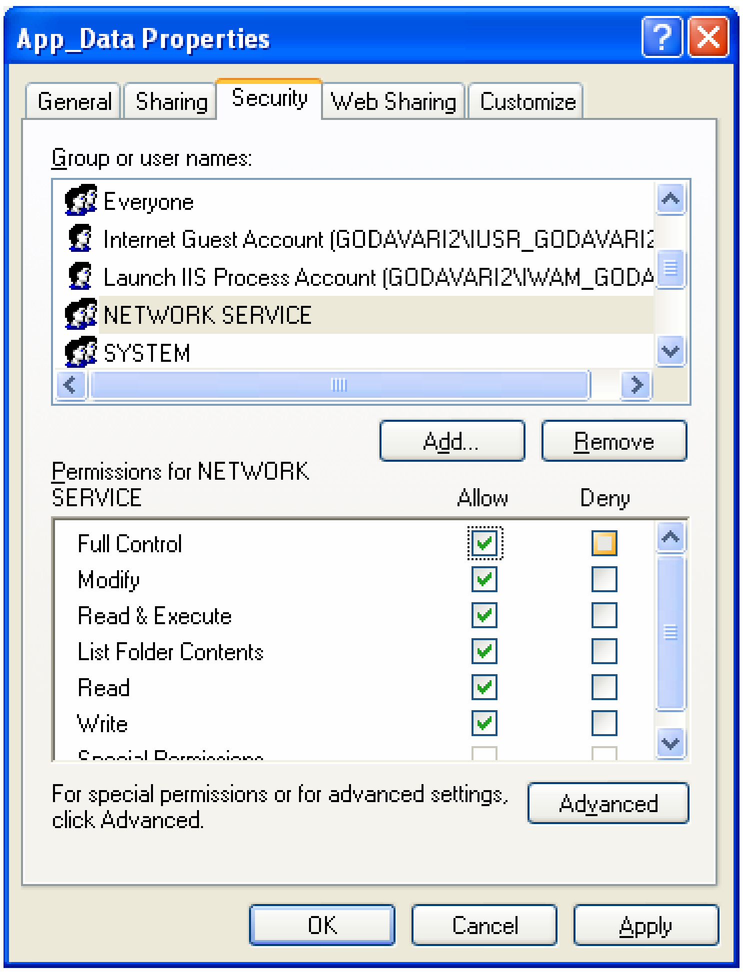
10. **Click** *OK* to confirm the user name.



11. **Set** the permissions, and then **click** *Add* if there are remaining user names that you need to add.



12. **Click** *OK* to close the Properties window for your folder.
13. Within your web service folder, navigate to the App_Data folder.
14. In the same manner as above, set permissions for the App_Data folder as follows:
    a. Network_Service - Full Control

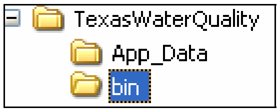


15. **Click** *OK* to close the Properties window.

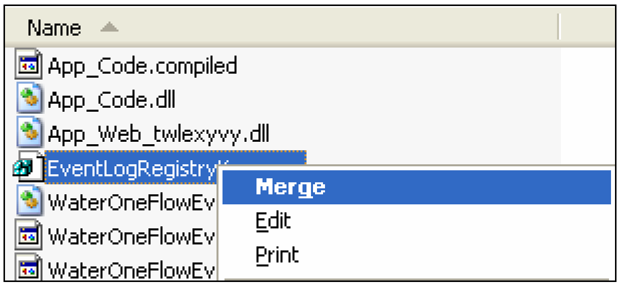
Register the Service

You will add a snippet of code to the Windows registry to let the computer know that your service is there.

1. In Windows Explorer, navigate to the folder for your web service (e.g. “TexasWaterQuality”), and then navigate within the *bin* folder.



2. In the *bin* folder, **right click** *EventLogRegistryKey.reg*, and **click** *Merge*.

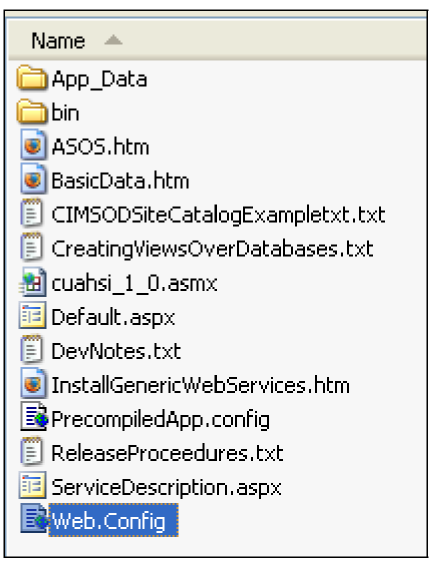


3. Accept any prompts asking for confirmation of the registry edit.

Edit Web.Config

You'll now edit the Web.Config for the service to give your service a network name and vocabulary, and specify the connection string to your database.

1. In the folder for your web service, locate the Web.Config file, and edit it.



Note: I've used Visual Studio 2005 as my editor in these screenshots.

2. In Web.Config, set the *network* and *vocabulary* for your service by **editing** the appropriate <add> elements. The network refers to the observation network to which your database pertains, such as "NWIS Daily Streamflow". The

vocabulary refers to the framework within which the terms that you use apply, such as “NWIS”. You can leave these as the default value of “ODM” if you wish.

```
            <add key="network" value="ODM"/>
            <add key="vocabulary" value="ODM"/>
```

3. **Edit** the connection string for your database. To find the connection string, location the <connectionStrings> element. Then find the <add> element that is a child of <connectionStrings>, which has a name of “ODDB”. This element has an attribute called “connectionString”, which you will edit as follows:
    a. If database is inside the App_data folder:

```
<add name="ODDB"
connectionString="Data Source=.\SQLEXPRESS;
Integrated Security=true;
User Instance=true;
AttachDbFilename=|DataDirectory|[FILENAME];
Database=[DATABSE_NAME]" providerName="System.Data.SqlClient"/>
```

    b. If database is a local file:

```
<add name="ODDB" connectionString="Data Source=.\SQLEXPRESS;
Integrated Security=true;
User Instance=true;
AttachDbFilename=[PATH_TO_FILE][FILENAME];
Database=[DATABSE_NAME]" providerName="System.Data.SqlClient"/>
```

    c. If database is a local database:

```
<add name="ODDB" connectionString="Data Source=.\SQLEXPRESS;
Integrated Security=true;
Database=[DATABSE_NAME]" providerName="System.Data.SqlClient"/>
```

    d. If database is remote:

```
<add name="ODDB" connectionString="Data Source=[HOSTNAME];
Integrated Security=true;
Database=[DATABSE_NAME]" providerName="System.Data.SqlClient"/>
```

Note: In the connection string, replace the items in brackets with your actual items (do not include the brackets):

- **[HOSTNAME] – the computer’s name hosting the database, e.g., DatabaseComputer**
- **[DATABASE_NAME] – e.g., MyOD**
- **[PATH_TO_FILE] – e.g., c:\Databases**
- **[FILENAME] – e.g., MyDatabase.mdf**

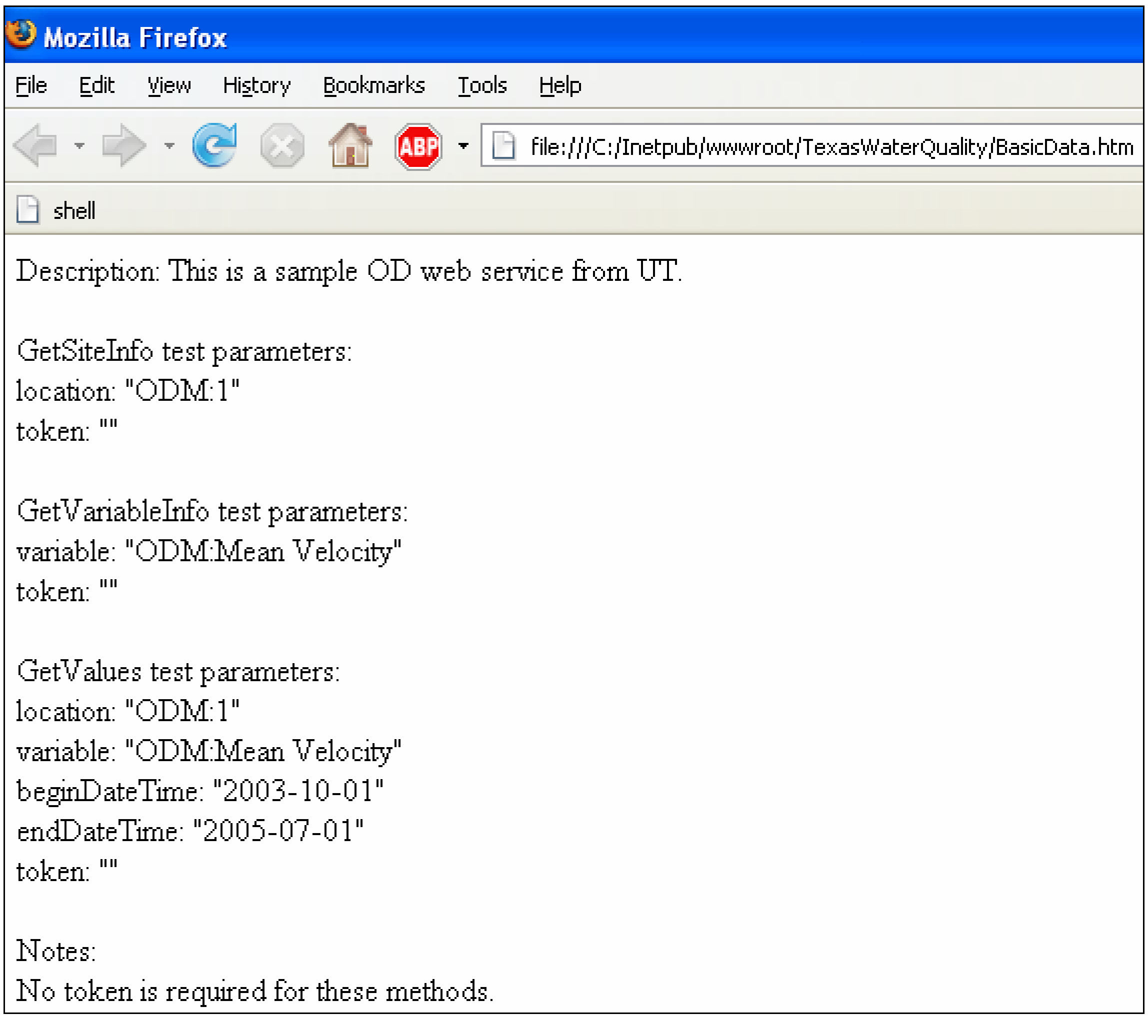
4. Save and close Web.Config

Edit BasicData.htm

BasicData.htm is appended to the web service description page when that page is viewed with an Internet browser. BasicData.htm gives information about your service, such as example parameters that can be used for testing the service. Edit this page to describe your service and data.

1. Edit BasicData.htm. You may put whatever content you like in this file.
2. Save and close BasicData.htm.

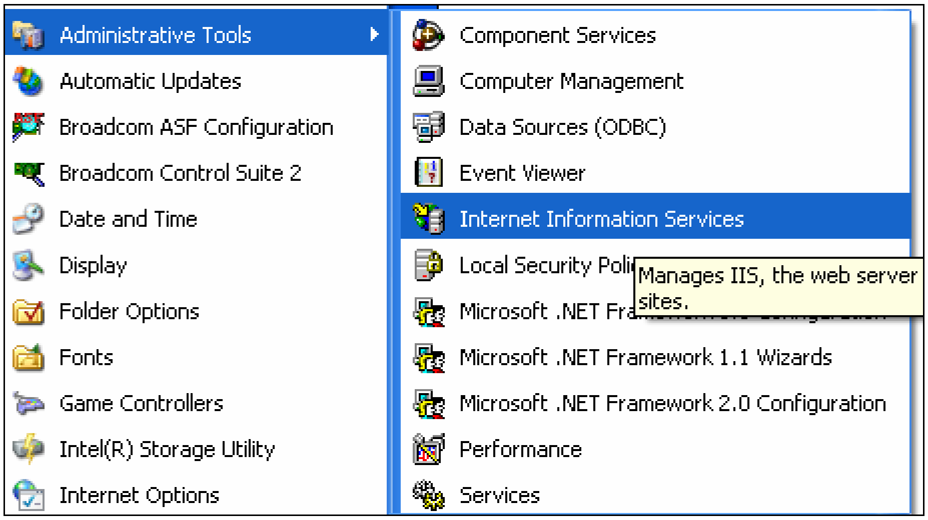
As an example, I edited BasicData.htm to show how to retrieve data from the sample OD.mdf database included with GenericODws.zip.



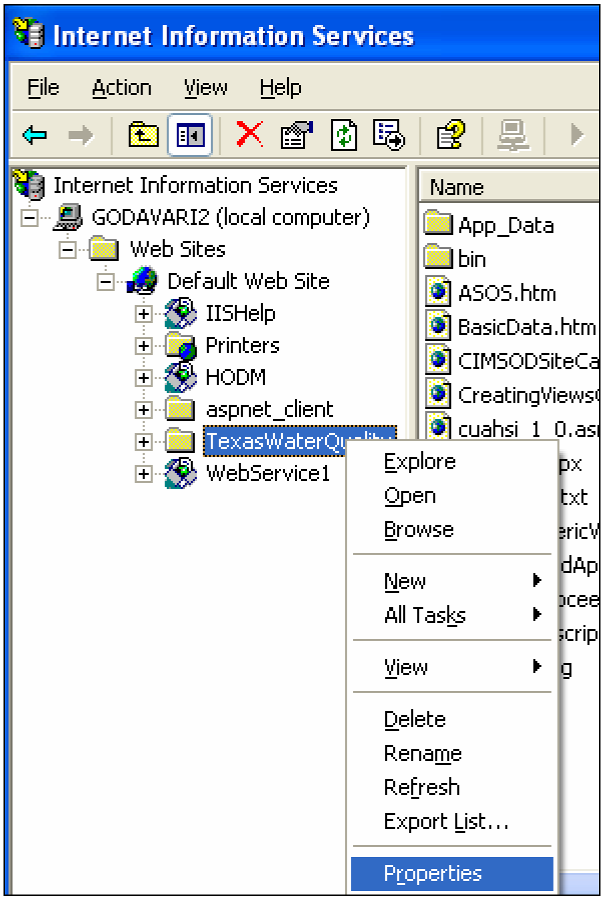
Configure Web Application in IIS Manager

In IIS Manager, you will now create a

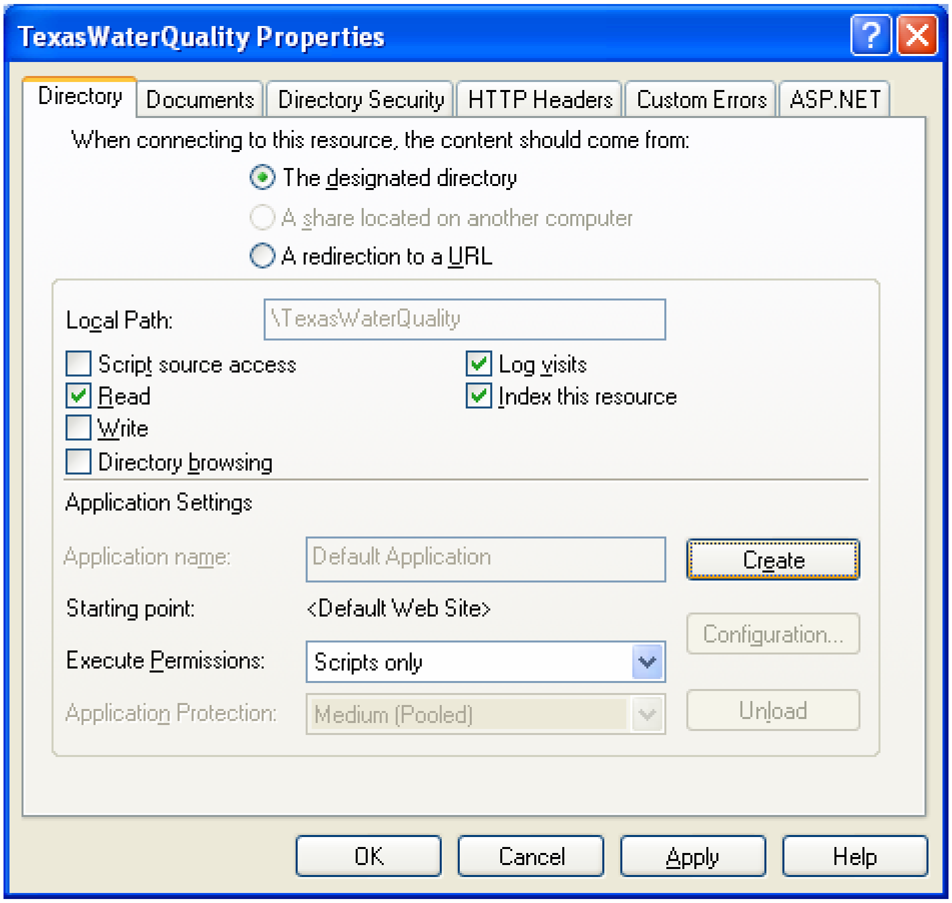
1. Open IIS Manager by **clicking** *Start→Control Panel→Administrative Tools→Internet Information Services*.



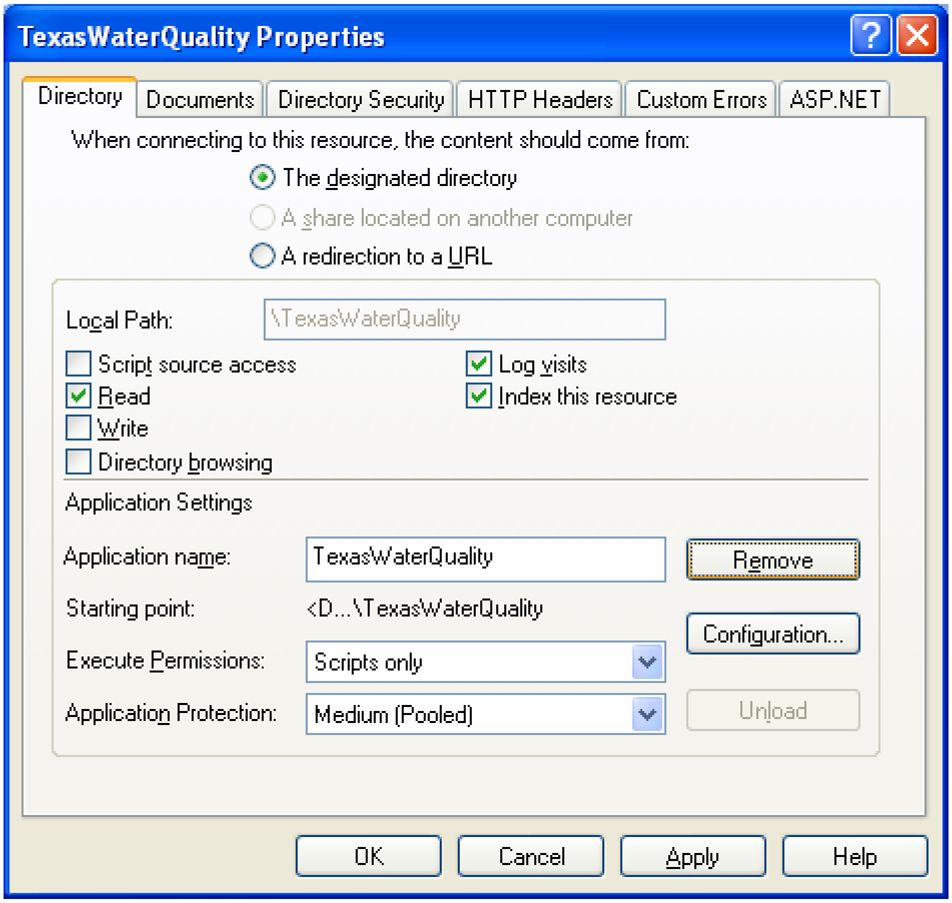
2. **Navigate** to *Web Sites→Default Web Site→[Your Web Service Folder Name]*.
3. **Right click** on your folder and **click** *Properties*.



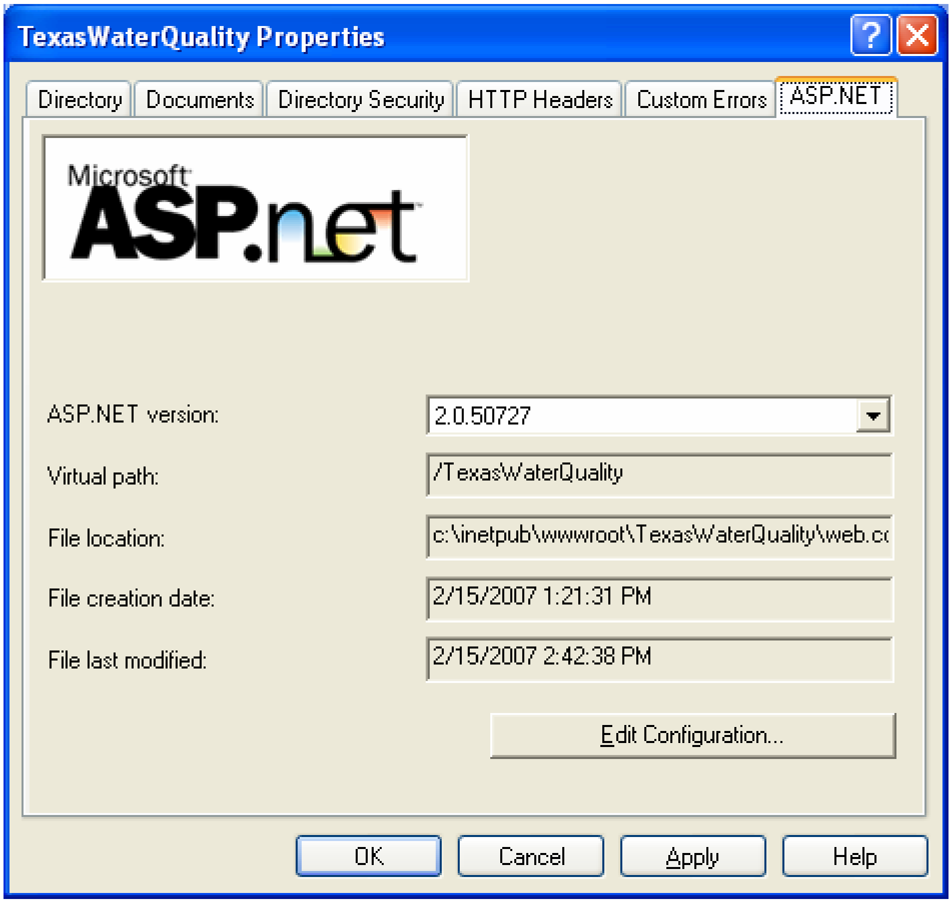
4. Click the *Directory* tab.



5. **Click** *Create*.



6. **Click** the *ASP.NET* tab. Make sure the ASP.NET version is 2 or higher.



7. **Click** *OK*.
8. **Close** IIS.

Test the Application

Now you're ready to test the application. Open a browser, and navigate to your localhost\[Service] folder, and invoke some calls for testing.

1. **Open** an Internet browser.
2. In the address bar of the browser, **enter** the following address:

http://localhost/[FOLDER NAME FOR YOUR SERVICE]/

3. Input test parameters and see if your service works.

Appendix C: NHDH Editing

C.1: Statistics for Subregions in Texas

Sub-region HUC4 Code	Subregion Name	Total # of Flowlines	Length of Flowlines (km)	Total Area (km^2)	# of Named Segments	Critical Discon-nects*
1108	Upper Canadian	48,582	37,513	32684	7873	0
1109	Lower Canadian	70,140	40,496	44019	12886	2
1110	North Canadian	69,835	42,590	45,996	13,610	5
1112	Red Headwaters	54,404	32,848	38,388	10,245	3
1113	Red-Washita	73,383	40,677	64,107	19,037	1
1114	Red-Sulphur	165,486	112,826	71,644	41,530	1
1201	Sabine	87,780	45,785	25,511	20,214	2
1202	Neches	86,836	44,399	25,780	20,140	1
1203	Trinity	156,968	76,426	46,572	32,860	2
1204	Galveston Bay-San Jacinto	45,780	29,351	28,069	8,045	1
1205	Brazos Headwaters	26,943	16,572	37,772	5,199	0
1206	Middle Brazos	136,334	61,400	40,297	27,390	0
1207	Lower Brazos	121,690	59,795	40,161	25,436	0
1208	Upper Colorado	22,779	16,705	41,426	4,571	2
1209	Lower Colorado-San Bernard Coastal	165,595	92,307	74,369	41,661	3
1210	Central Texas Coastal	105,059	60726.769	51,434	24,612	0
1211	Nueces-Southwestern Texas Coastal	106,723	70840.067	82,596	22,522	1
1303	Rio Grande-Mimbres	45,301	35388.080	38,733	2,664	0
1304	Rio Grande-Amistad	89,025	77720.351	86,893	7,648	4
1305	Rio Grande Closed Basins	48,937	47129.208	45,644	1,576	15+
1306	Upper Pecos	52,112	45623.960	61,165	4,988	2
1307	Lower Pecos	43,187	39491.009	53,618	2,609	4
1308	Rio Grande-Falcon	27,929	17138.080	29,257	4,889	0
1309	Lower Rio Grande	716	1858.899	16,627	243	1
	Total	**1,802,942**	**1,108,094**	**1,090,080**	**354575**	

*The data used for this analysis were downloaded in September, 2005. Some of the errors discovered with this data have since been fixed by the USGS. Thus, the number of Critical Disconnects in this analysis is greater than the number of disconnects actually fixed for this project.

C.2: DATA SOURCES USED

Data Source	Location
NHD 24k Original	Huan.sm_work.SDE.Hydro_NHDnGEO_9_06. NHDFLOWLINE_TX_SX
NHD 24k Edited	Huan.sm_work.SDE.NHDFLOWLINE_TJ_EDIT
Digital Orthophoto Quarter Quad aerial Imagery	Elrond.tnris_doqqs.sde.TX_DOQQS_Z13, Z14, Z15
USGS Digital Raster Graphic Maps	Elrond.stratmapraster.sde.DRG_24kZ13, Z14, Z15
NHD 24k- geodatabase with network	ftp://nhdftp.usgs.gov/SubRegions/High/
National Elevation Dataset	Elrond.stratmapraster.sde.NED_Dec_04
ESRI Geography Network Satellite Imagery	http://www.geographynetwork.com/explorer/

C.3: FIXES MADE

The following is a list of COMID's of disconnects fixed. Some fixes required the adjustment of multiple flowlines. Thus the number of COMID's actually adjusted is greater than the following list, and is reflected in the NHDFeatureToMetadata table.

COMID
129311673
95111908
95111530
94993689
92590420
130012590
108565635
137785150
137857006
106297264
125247405
136289623
137854505
129311813
109564499
131534552
101686349
140776101
65352624
133632137

C.4: NHDMETADATA AND NHDFEATURETOMETADATA TABLES

NHDMetadata

DUUID	Process Description	Process Date
90001	Extended or moved reach to increase connectivity and to reflect physical flow	11/21/2006
90002	Created new reach to increase connectivity and to reflect physical flow	11/21/2006
90003	Changed FLOWDIR field from "Uninitialized" (0) to "With Digitized" (1)	11/21/2006
90004	Deleted duplicated flowline	11/21/2006
90005	Changed FLOWDIR field from "With Digitized" (1) to "Against Digitized" (2) to match 24k DRG, and to increase connectivity	12/7/2006

Note: Additional fields exist describing the agency responsible for the data, as well as the accuracy of the edits. See the actual table in the Geodatabase supplied with the deliverables.

NHDFeature to Metadata

OBJECTID	COMID	DUUID
1	129311813	90001
2	109564499	90001
3	131534552	90001
4	101686349	90001
5	92590420	90001
6	95111530	90001
7	95111908	90001
8	137785150	90001
9	137857006	90001
10	106297264	90001
11	133632137	90003
12	65352624	90004
13	140776101	90001
14	136289623	90001
15	94993689	90001
16	200000001	90002
17	108537899	90001
18	108537899	90005
19	108565635	90001
20	108537897	90005
21	108565633	90005

OBJECTID	COMID	DUUID
22	108565635	90005
23	108537897	90001
24	137854505	90001
25	200000002	90002
26	200000003	90002
27	200000004	90002
28	200000005	90002
29	129311679	90001

Appendix D: Conversion of SWQM to the Observations Data Model

D.1.A: SWQM EVENT TABLE DESCRIPTION (*TCEQ*, 2007A)

```
EVENT FILE

Tag_id     A7    This field is the key between the event and results
                 tables and is 7 characters long. The first character(s)
                 is the  prefix code
                 for the submitting agency.

Station    A9    This is a combination of the segment_id and the sequence
                 of a site within a segment

Stationid A5     This is a unique id that identifies each sampling
                 station.
                 This number is generated by the TNRCC.
                 Enddate   A10  The date the sample was collected in the
                 form of MM/DD/YYYY

Endtime    A5    The time the sample was collected in military format
                 (HH:MM)

Enddepth   A6    This is the depth in meters at which the sample was
                 collected.

Startdate A10    This field is only required for composite samples and is
                 the beginning date in the form of MM/DD/YYYY

Starttime A5     This field is only required for composite samples and is
                 the beginning time (in military format) at which the
                 sample was collected (HH:MM)

Startdepth A6    This field is only required for composite samples and is
                 the depth nearest surface (in meters) at which the
                 sample was collected.

Category   A1    This field is only required for composite samples and
                 should correspond to the following codes:
                         T    is for time composites
                         S    is for space composites (i.e.depth)
                         B    is for both space and time composites
                         F    is for flow weighted composites

Calculatn A1     This field is no longer used and should be left blank

Type       A2    This field is only required for composite samples and
                 should correspond to the following codes:
                         CN   for continuous
```

		## where ## is the number of grabs in the composite GB where the number of grabs is unknown
Comment	A135	This is a text field where record of any observational data is included with the sample
Source1	A2	The TNRCC assigned code for the submitting agency.
Source2	A2	An optional field that may be used to further identify the sample
Program	A2	A field that further identifies the sample. This field may be used to tie targeted monitoring to specific permits.

D.1.B: SWQM RESULT TABLE DESCRIPTION (*TCEQ*, 2007B)

RESULTS FILE

Tag_id	A7	This field is the key between the event and results tables and is 7 characters long. The first character(s) is the prefix code for the submitting agency.
Enddate	A10	The date the sample was collected in the form of MM/DD/YYYY
Storetcode	A5	This is a five digit code which identifies the substance or measurement.
Gtlt	A1	If the value is above the detection limit then this field should contain an >. If the value is below the detection limit then this field should contain an <.
Value	A8	This is the test result and should be reported in units according to the storet description

D.1.C: SWQM STATIONS TABLE DESCRIPTION (*TCEQ*, 2007C)

Basin Id	The number assigned to one of Texas' 24 River or Coastal Basins and the bays and estuaries
Station Id	The unique identifier assigned by the TNRCC to a sampling station
Station Num	The unique identifier assigned by the TNRCC placing it in sequence with all other sampling sites in a river basin. It is made up of river segment, a dot, and the sequence. For example : 0100.0105 means it is on segment 100 and its sequence number is .0105.
USGS Gage	A unique identifier assigned by the USGS to a gage station
Short Description	Text describing a sampling site's location
Long Description	Text describing a sampling site's location
EPA Type1	An abbreviated term that indicates where a sampling site is located. Examples: STREAM, RESERVOIR, PIPE
EPA Type2	An abbreviated term that indicates the conditions at a sampling site is located. Examples: AMBNT, TREATD, NTRTMT
County Name	The fullname of a county in which a station is located
County Id	The code assigned to a county by the state of Texas
Segment Id	A code assigned to a classified stream segment
Stream Sequence No.	A code assigned by TNRCC to a sampling site, placing it in with all other sampling sites in a river basin
Region	The TNRCC region in which a sampling site is located
Latitude	The latitude of the sampling site.
Longitude	The longitude of the sampling site
HUC	The eight digit hydrologic unit code assigned by the USGS towatersheds within each river basin
On Segment Flag	1 means it is on segment, 0 means it is not.

D.1.D: SWQM PARAMETER TABLE DESCRIPTION (*TCEQ*, 2007D)

Storet Code (A5)
Short Description 1 (A8)
Short Description 2 (A8)
Short Description 3 (A8)
Long Description (A50)
Minimum Value - minimum value allowed (Number)
Maximum Value - maximum value allowed (Number)

D.2.A: SWQM Event Table Example

Tag_id	Sta-tion	Station id	End date	End time	End depth	Start-date	Start-time	Start-depth	Cate-gory	Calc-ulatn	Ty-pe	Comment	Source 1	Source 2	Pro-gram
RR 10655	102 .01	10036	3/13/ 2000	7:30	0.3			0					RR	CR	RT
RR 10676	102 .01	10036	4/24/ 2000	8:35	0.3			0					RR	CR	RT
RR 10696	102 .01	10036	6/6/ 2000	8:30	0.3			0				Parameter code 00950 corrected to 00951 per CRP request on 10 Sep 2003.	RR	CR	RT
RR 10721	104 .01	10058	8/9/ 2000	12:30	0.3			0				DO measured extremely low; cleaned film on probe.	RR	RR	RT
31791	102 .01	10036	7/25/ 2000	13:20	0.1			0					WC	FO	RT
RR 10743	104 .01	10058	12/4/ 2000	10:30	0.3			0				Site has had significant rainfall in the past 30 days.	RR	RR	RT
31792	102 .01	10036	7/25/ 2000	14:55	0.3			0					WC	FO	RT
31794	102 .01	10036	7/25/ 2000	15:15	25.48	7/25 /2000	15:00	25.48	T		3		WC	FO	RT
R1 92668	102 .01	10036	7/25/ 2000	15:20	0.3			0				SKY: PARTLY	WC	FO	RT

Tag_id	Station	Station id	End date	End time	End depth	Start-date	Start-time	Start-depth	Category	Calculatn	Type	Comment	Source 1	Source 2	Program
												CLOUDY; TEMP: 90 F; WIND: SE 5 MPH; WATER COLOR: CLEAR; LIGHT REC. ACTIVITY; CALM DAY; VERY CLEAR WATER; LAKE OFFICIALLY AT			
PR 32671	102.01	10036	7/25/2000	15:20	6.1			0					WC	FO	RT
RR 10710	104.01	10058	6/12/2000	12:55	0.3			0				No Bact due to rain w/in 24 hrs; No metals due to wtr too turbid	RR	RR	RT
PR 32678	102.01	10036	7/25/2000	15:20	24.99			0					WC	FO	RT
RR 10767	102.01	10036	8/22/2000	8:07	0.3			0					RR	CR	RT
H0 50000	102.01	10036	10/3/2000	16:00	3			0					HD	HD	SS

D.2.B: SWQM RESULT TABLE EXAMPLE

Tag_id	Enddate	Storetcode	Gtlt	Value
27418	1/10/2000	410		272
27418	1/10/2000	530		20
27418	1/10/2000	535		6
27418	1/10/2000	593		3.12
27418	1/10/2000	610		0.71
27418	1/10/2000	625		1.27
27418	1/10/2000	665		0.35
27418	1/10/2000	680		10
27418	1/10/2000	940		1330
27418	1/10/2000	945		569
27418	1/10/2000	32211		7.21
27418	1/10/2000	32218		2.32
27418	1/10/2000	70300		4210
27418	1/10/2000	70507	<	0.06
27419	1/10/2000	915		265
27419	1/10/2000	925		111
27419	1/10/2000	1000		2.67
27419	1/10/2000	1025	<	4
27419	1/10/2000	1030	<	3
27419	1/10/2000	1040	<	3
27419	1/10/2000	1049	<	2
27419	1/10/2000	1065	<	10
27419	1/10/2000	1075	<	0.25
27419	1/10/2000	1090	<	8
27419	1/10/2000	1106	<	25
27419	1/10/2000	1145	<	10

D.2.B: SWQM STATIONS TABLE EXAMPLE

Basin Id	Station Id	Station Num	USGS Gage Station	Short Description	Long Description	EPA Type 1	EPA Type 2	County Name	County Id	Segment Id	Stream Sequence No.	Region	Latitude	Longitude	HUC	On Segment Flag
1	10002	100.0105		LAKE MARVIN AT MIDLAKE	LAKE MARVIN AT MIDLAKE 10 MI. EAST OF CANADIAN	RESERV	AMBNT	Hemphill	106	101	0	1	35.889999	-100.192497	11090106	0
1	10003	100.02	7227448	PUNTA DE AGUA CREEK FM 767	PUNTA DE AGUA CREEK AT FM 767 WEST OF CHANNING	STREAM	AMBNT	Hartley	103	103	0	1	35.668056	-102.481392	11090101	0
1	10004	100.03		COLD WATER CREEK AT SH 136	COLDWATER CREEK AT SH 136 NORTH OF GRUVER	STREAM	AMBNT	HANSFORD	98	199	0	1	36.433334	-101.451668	11100103	0
1	10005	100.038		PALO DURO RESERVOIR NEAR DAM	PALO DURO RESEVOIR AT BOAT LAUNCH NEAR DAM 19 KM NORTH OF SPEARMAN	RESERV	AMBNT	HANSFORD	98	199	0	1	36.360001	-101.164444	11100103	0
1	10006	100.04		PALO DURO	PALO DURO CREEK AT	STREAM	AMBNT	HANSFORD	98	199	0	1	36.2733	-101.	1110	0

Basin Id	Station Id	Station Num	USGS Gage Station	Short Description	Long Description	EPA Type 1	EPA Type 2	County Name	County Id	Segment Id	Stream Sequence No.	Region	Latitude	Longitude	HUC	On Segment Flag
				CREEK AT FM 2387	FM 2387 NORTHWEST OF SPEARMAN								35	254997	0103	
1	10007	100.042	7233500	PALO DURO CREEK AT SH 207-15	PALO DURO CREEK SH 207-15 WEST OF SPEARMAN	STREAM	AMBNT	HANS-FORD	98	199	0	1	36.202221	-101.305557	11100103	0
1	10008	100.05		SOUTH PALO DURO CRK AT SH 136	SOUTH PALO DURO CREEK AT SH 136 NORTH OF PRINGLE	STREAM	AMBNT	Hutch-inson	117	199	0	1	36.006668	-101.463333	11090106	0
1	10009	100.06		KIOWA CREEK AT SH 15	KIOWA CREEK AT SH 15 EAST OF DARROUZETT	STREAM	AMBNT	LIPS-COMB	148	199	0	1	36.446667	-100.308334	11100201	0

D.2.B: SWQM PARAMETERS TABLE EXAMPLE

Storet Code	Short Description 1	Short Description 2	Short Description 3	Long Description	Minimum Value	Maximum Value
3	SAMPLOC	DEPTH	FEET	SAMPLING STATION LOCATION, VERTICAL FEET	0	300
4	STREAM	WIDTH	FEET	STREAM WIDTH (FEET)	0.1	700
10	WATER	TEMP	CENT	TEMPERATURE, WATER (DEGREES CENTIGRADE)	1	38
11	WATER	TEMP	FAHN	TEMPERATURE, WATER (DEGREES FAHRENHEIT)	32	99
20	AIR	TEMP	CENT	TEMPERATURE, AIR (DEGREES CENTIGRADE)	-15	45
21	AIR	TEMP	FAHN	TEMPERATURE, AIR (DEGREES FAHRENHEIT)	-55555556	110
23	WEIGHT		POUNDS	SAMPLE WEIGHT IN POUNDS	0.01	100
24	LENGTH		INCHES	SAMPLE LENGTH IN INCHES	0.01	100
29	SAMPLE	SEQUENCE	NUMBER	FIELD SPECIMEN OR SAMPLE SEQUENCE NUMBER	-55555556	55555556
55	STREAM	VELOCITY	FT/SEC	STREAM VELOCITY (FEET PER SECOND)	0	10
56	FLOW	RATE	GPD	FLOW RATE (GALLONS PER DAY)	0	9999999
60	STREAM	FLOW	CFS	FLOW, STREAM, MEAN DAILY (CUBIC FEET PER SEC)	0	99999
39	SAMPLE	LENGTH	MM	SAMPLE LENGTH IN MILLIMETERS	0.3	2540
19	SAMPLE	WEIGHT	GRAMS	SAMPLE WEIGHT IN GRAMS	299	1000
54	RESVOIR	STORAGE	ACRE FT	RESERVOIR STORAGE - ACRE FEET	-55555556	55555556
61	STREAM	FLOW	INST-CFS	FLOW STREAM, INSTANTANEOUS (CUBIC FEET PER SEC)	0.01	15000
62	WATER	SURF ELE	IN FEET	ELEVATION, RESERVOIR SURFACE WATER IN FEET	1	3000
64	DEPTH OF	STREAM	MEAN(FT)	DEPTH OF STREAM, MEAN (FT)	0.01	300
65	STREAM	STAGE	FEET	STAGE, STREAM (FEET)	1	3000
68	MAX SAMP	DEPTH	FEET	DEPTH, MAXIMUM, OF SAMPLE (FEET)	0.1	300
70	TURB	JKSN	JTU	TURBIDITY, (JACKSON CANDLE UNITS)	0	40
76	TURB	TRBIDMTR	HACH FTU	TURBIDITY,HACH TURBIDIMETER (FORMAZIN TURB UNIT)	0	5000

D.3: SWQM TO ODM FIELD MAP

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
Event	Tag_id	GroupDescriptions	GroupDescription	Each group description is "Tag_id = xxxxx". These are linked back to the value (result) via GroupID and the Groups table.
Event	Comment	Qualifiers	QualifierDescription	Alternatively, these "comments" could be made into their own values (where they apply). Or, could add to GroupDescription, along w/ Tag_id
Event	Category	Samples	SampleType	Alternatively, use Methods:MethodDescription. Either way, multiple values of [SampleType, MethodDescription] will be used, one for each combination of Category and Type
Event	Type	Samples	SampleType	Alternatively, use Methods:MethodDescription. Either way, multiple values of [SampleType, MethodDescription] will be used, one for each combination of Category and Type
Event	Source1	Sources	Organization	the "Organizatation" field will be filled with information on the original source (I.e. not SWQM)- am trying to get more info than just the codes we have now
Event	Source2	Sources	Organization	
Event	Program	Sources	Organization	
Event	Stationid	Values	SiteCode	
Event	Enddate	Values	DateTime	Add to Event:Endtime
Event	Endtime	Values	DateTime	Add to Event:Enddate
Event	Enddepth	Values	OffsetValue	
Event	Station			No Match in ODM4
Event	StartDate			No Match in ODM4
Event	StartTime			No Match in ODM4
Event	StartDepth			No Match in ODM4
Event	Calculatn			No Match in ODM4
Parameter	Storet Code	Variables	VariableCode	
Parameter	Long	Variables	VariableName	

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
	Description			
Parameter	Minimum Value			No Match in ODM4
Parameter	Maximum Value			No Match in ODM4
Parameter	Short Description 1			No Match in ODM4
Parameter	Short Description 2			No Match in ODM4
Parameter	Short Description 3			No Match in ODM4
Result	Gtlt	Values	CensorCode	If <, 'lt'; >, 'gt'
Result	Value	Values	Value	
Result	Storetcode	Variables	VariableCode	
Result	Tag_id			See Event:Tag_id
Result	Enddate			No Match in ODM4
Stations	HUC	Sites	Comments	"HUC 8 = xxxxxxxx"
Stations	County Name	Sites	County	
Stations	Latitude	Sites	Latitude	
Stations	Longitude	Sites	Longitude	
Stations	Station Id	Sites	SiteCode	
Stations	Long Description	Sites	SiteName	
Stations	USGS Gage Station			No Match in ODM4
Stations	EPA Type1	Sites	Comments	"EPA Type1 =xxxxx"
Stations	EPA	Sites	Comments	"EPA Type2 = xxxxx"

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
	Type2			
Stations	Basin_Id			No Match in ODM4
Stations	Station Num			No Match in ODM4
Stations	Short Description			No Match in ODM4
Stations	County Id			No Match in ODM4
Stations	Segment Id			No Match in ODM4
Stations	Stream Sequence No.			No Match in ODM4
Stations	Region			No Match in ODM4
Stations	On Segment Flag			No Match in ODM4
		Categories	CategoryDescription	No Match in SWQM
		Categories	VariableID	No Match in SWQM
		Categories	Value	No Match in SWQM
		DerivedFrom	DerivedFromID	No Match in SWQM
		DerivedFrom	ValueID	No Match in SWQM
		Groups	GroupID	Link GroupDescriptions (filled with Tag_id from Events) to Values (each Result has a Tag_id)
		Groups	ValueID	Link GroupDescriptions (filled with Tag_id from Events) to Values (each Result has a Tag_id)
		ISOMetaData	MetaDataID	No Match in SWQM
		ISOMetaData	TopicCategory	No Match in SWQM
		ISOMetaData	Title	No Match in SWQM
		ISOMetaData	Abstract	No Match in SWQM
		ISOMetaData	ProfileVersion	No Match in SWQM
		ISOMetaData	MetadataLink	No Match in SWQM
		LabMethods	LabMethodID	No Match in SWQM

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
		LabMethods	LabName	No Match in SWQM
		LabMethods	LabOrganization	No Match in SWQM
		LabMethods	LabMethodName	No Match in SWQM
		LabMethods	LabMethodDescription	No Match in SWQM
		LabMethods	LabMethodLink	No Match in SWQM
		Methods	MethodID	No Match in SWQM
		Methods	MethodDescription	No Match in SWQM
		Methods	MethodLink	No Match in SWQM
		OffsetTypes	OffsetTypeID	Link to Events:EndDepth
		OffsetTypes	OffsetUnitsID	"52"
		OffsetTypes	OffsetDescription	"Depth below water surface level"
		Qualifiers	QualifierID	Link Events:Comments to a specific Record
		Qualifiers	QualifierCode	No Match in SWQM
		Quality-ControlLevels	Definition	"This data has been subjected to a limited quality control check."
		Quality-ControlLevels	QualityControlLevel	"2"
		Quality-ControlLevels	Explanation	"This data has been visually quality control checked. No systematic QC algorithms have been performed."
		Samples	SampleID	A unique SampleID for each combination of Event:Category and Event:Type is created, linking SampleType to Value
		Samples	LabSampleCode	No Match in SWQM
		Samples	LabMethodID	No Match in SWQM
		SeriesCatalog	SeriesID	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	SiteID	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	SiteCode	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	SiteName	Generated after data is loaded using tool created by Jeff Horsburgh at USU

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
		SeriesCatalog	VariableID	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	VariableCode	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	VariableName	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	VariableUnitsID	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	VariableUnitsName	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	SampleMedium	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	ValueType	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	BeginDateTime	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	EndDateTime	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	UTCOffset	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	ValueCount	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		SeriesCatalog	GeneralCategory	Generated after data is loaded using tool created by Jeff Horsburgh at USU
		Sites	Elevation_m	No Match in SWQM
		Sites	VerticalDatum	No Match in SWQM
		Sites	LatLongDatumID	"2" Link to Spatial References
		Sites	PosAccuracy_m	No Match in SWQM
		Sites	SiteID	Generated by ODM
		Sites	LocalX	No Match in SWQM
		Sites	LocalY	No Match in SWQM
		Sites	LocalProjectionID	No Match in SWQM
		Sites	State	"Texas"

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
		Sources	ContactName	
		Sources	Email	crp@tceq.state.tx.us <crp@tceq.state.tx.us>
		Sources	Address	TCEQ, Contact Name, Mail Code, P.O. Box 13087
		Sources	MetaDataID	No Match in SWQM
		Sources	SourceID	Generated by ODM
		Sources	SourceDescription	"Text file retrieved from TRACS SWQM program, with data originally from numerous public and private monitoring organizations. See http://www.tceq.state.tx.us/assets/public/compliance /monops/water/wdma/dmrg/2005/2005dmrg_complete.pdf for more information about the source."
		Sources	SourceLink	http://www.tceq.state.tx.us/compliance /monitoring/crp/data/samplequery.html
		Sources	City	"Austin"
		Sources	State	"TX"
		Sources	ZipCode	"78711-3087"
		SpatialReferences	SpatialReferenceID	"2"
		SpatialReferences	SRSID	"4269"
		SpatialReferences	SRSName	"NAD83"
		SpatialReferences	IsGeographic	TRUE
		SpatialReferences	Notes	
		Units	UnitsID	Mapping units from the Parameter table would be extremely time consuming. For the moment, the Variables Description includes units, and will suffice.
		Units	UnitsName	Mapping units from the Parameter table would be extremely time consuming. For the moment, the Variables Description includes units, and will suffice.
		Units	UnitsType	Mapping units from the Parameter table would be extremely time consuming. For the moment, the Variables Description includes units, and will suffice.
		Units	UnitsAbbreviation	Mapping units from the Parameter table would be extremely time consuming. For the moment, the Variables Description includes units, and will suffice.

SWQM Table	SWQM Field	ODM4 Table	ODM4 Field	Comments
		Values	QualityControlLevel	2
		Values	SiteID	Link to Sites Table
		Values	SourceID	Link to Sources Table
		Values	VariableID	Link to Variables Table
		Values	AccuracyStdDev	No Match in SWQM
		Values	OffsetTypeID	Link to OffsetTypes Table
		Values	ValueID	Generated by ODM
		Values	UTCOffset	-6
		Values	UTCDateTime	=Values:DateTime - 6
		Values	QualifierID	Link to Qualifiers table- links Value to a comment in Qualifier Description
		Values	MethodID	No Match in SWQM
		Values	SampleID	Link to SampleType, which includes Event:Category and Event:Type
		Values	DerivedFromID	No Match in SWQM
		Variables	TimeSupport	No Match in SWQM
		Variables	ValueType	"Field Observation"
		Variables	VariableID	Generated by ODM
		Variables	VariablesUnitsID	No Match in SWQM
		Variables	TimeUnitsID	No Match in SWQM
		Variables	DataType	No Match in SWQM
		Variables	SampleMedium	"Surface Water"
		Variables	IsRegular	No Match in SWQM
		Variables	GeneralCategory	"Water Quality"
		Variables	NoDataValue	No Match in SWQM

D.4: SWQM TO ODM SQL SERVER INTEGRATION SERVICES SCRIPT

```
-- The following SQL Server script was used for the specific purpose of
migrating TRACS SWQM data into the ODM v 4.0.  It is to be used as an
instructional example of how to use SQL Server scripts.  It will not
work as written with other ODM 4 loading operations.  Major edits will
need to be made, including changing the names of tables, databases and
specific fields to be loaded.  This script was written by Chris
Williams at the Texas Natural Resources Information System (TNRIS) and
Tyler Jantzen at the University of Texas at Austin.


>>>>>>
use odm_tracs
go

--example
/*
INSERT INTO MyTable  (PriKey, Description)
       SELECT ForeignKey, Description
       FROM SomeView
*/

--Template
/*
--***************
begin tran

Select    *
from

insert    into       ()
          Select
          From      dbo.TRACS_

Select    *
from

rollback tran
--commit tran
*/

--"rollback" and "commit" tran can be turned on and off by commenting
out
--"rollback" runs the transaction, but does not actually store the
results to
-- the database.  "commit" runs the transaction and stores the results
to the
-- database.
-- only one of the two functions should be activated

-- To actually run the entire script (and store results),
```

```
-- comment out "rollback", and activate "commit"

-- The following order of loading data into the ODM corresponds with
the
-- order described in Section 4.4.3.3 and Table 4.6

-- The "#" sign indicates the creation of a tempoarary table.
-- Temporary tables were created for some transfer operations
-- where an intermediate table was necessary


--***************GroupDescription
begin tran

Select    *
from      groupdescriptions

insert    into groupdescriptions (GroupDescription)
          Select Tag_id
          From dbo.Tracs_event

Select    *
from      groupdescriptions

rollback tran
--commit tran


--***************ISOMetadata--already predefined as part of ODM.  No
match in SWQM.

--***************LabMethods - No match in SWQM!!!

--***************Methods

--***************Qualifiers

begin tran

select *
from qualifiers

Select Distinct (comment) as comment
into #tracsevent_comment
from dbo.tracs_event

delete
from #tracsevent_comment
where comment is null or comment = ''

select *
from #tracsevent_comment
```

```
insert    into qualifiers (qualifierdescription)
          Select (comment)
          from #tracsevent_comment

select *
from qualifiers

drop table #tracsevent_comment

rollback tran
--commit tran


--***************QualityControlLevels - No Match in SWQM!

--***************SPATIAL REFERENCES
begin tran

Select    *
from      SpatialReferences

insert    into      spatialreferences(SRSID,SRSName,IsGeographic,Notes)
          Values    ('4269', 'NAD83', 1 , '')

Select    *
from      spatialreferences

rollback tran
--commit tran

--***************Units - No Match in SWQM (the units are part of
Variables description)

--***************OffsetTypes

begin tran

Select    *
from      OffsetTypes

insert    into OffsetTypes (OffsetUnitsID, OffsetDescription)
          Values  (52, 'Depth below water surface level')
-- The OffsetUnitsID "52" assumes that the units table from the ODM4
example as loaded,
-- and that the 52nd record is "meters".  If this is not the case, the
unit "meters"
-- needs to be loaded into the ODM, and the corresponding UnitsID
replace the value "52" above

Select    *
from

rollback tran
```

```
--commit tran
*/


--**************Samples
begin tran

Select    *
from      dbo.samples

 _
select category + '_' + [type] as cattype
into #tracsevent_cattype
from dbo.tracs_event

delete
from #tracsevent_cattype
where cattype is null or cattype = '' or cattype = ' _  '

insert    into      samples (sampletype)
          Select    distinct (cattype)
          From      #tracsevent_cattype


select distinct (cattype)
from #tracsevent_cattype

Select    *
from      dbo.samples

drop table #tracsevent_cattype

rollback tran
--commit tran


--**************SITES
begin tran

Select    *
from      Sites

Select    Station_ID, Long_Desc, Latitude, Longitude, County_Name, 'HUC
8 = ' + HUC + ';' + ' EPA_Type1 = ' + EPA_Type1 + '; EPA_Type2 = ' +
EPA_Type2 as Comments
into      #Sites_values
from      dbo.Tracs_Stations


Alter Table #Sites_Values ADD [State] char(10) Null;
go
Alter Table #Sites_Values ADD [LatLongDatumID] char(2) null;
go
```

```
update    #Sites_Values
Set       [State] = 'Texas',
          LatLongDatumID = '2'

Select    *
from      #Sites_Values

insert    into      Sites     (SiteCode,
                               SiteName,
                               Latitude,
                               Longitude,
                               LatLongDatumID,
                               [State],
                               County,
                               Comments)
          Select
                    Station_ID,
                    Long_Desc,
                    Latitude,
                    Longitude,
                    LatLongDatumID,
                    [State],
                    County_Name,
                    Comments
          From      #Sites_Values

Select    *
from      Sites

drop table #sites_Values

rollback tran
--commit tran

--**************Sources
begin tran

Select    *
from      Sources

Drop table #Sources_Dist

Select    distinct source1, source2, program
into      #Sources_Dist
from      dbo.Tracs_Event

select    *
from      #Sources_Dist

Alter Table #Sources_Dist ADD [Organization] char(255) Null;
go
Alter Table #Sources_Dist ADD [SourceDescription] char(255) Null;
```

```
go
Alter Table #Sources_Dist ADD [SourceLink] char(255) Null;
go
Alter Table #Sources_Dist ADD [ContactName] char(50) Null;
go
Alter Table #Sources_Dist ADD [Phone] char(50) Null;
go
Alter Table #Sources_Dist ADD [Email] char(50) Null;
go
Alter Table #Sources_Dist ADD [Address] char(255) Null;
go
Alter Table #Sources_Dist ADD [City] char(50) Null;
go
Alter Table #Sources_Dist ADD [State] char(50) Null;
go
Alter Table #Sources_Dist ADD [ZipCode] char(50) Null;
go

update     #Sources_Dist
Set        Organization = 'TCEQ Sources = ' + Source1 + '; ' + Source2 +
'; ' + Program,
           SourceDescription = 'Text file retrieved from TRACS SWQM
program, with data originally from numerous public and private
monitoring organizations.  See
http://www.tceq.state.tx.us/assets/public/compliance/monops/water/wdma/
dmrg/2005/2005dmrg_complete.pdf for more information.',
           SourceLink =
'http://www.tceq.state.tx.us/compliance/monitoring/crp/data/samplequery
.html',
           ContactName = '',
           Phone = '512-239-3282',
           Email = 'crp@tceq.state.tx.us',
           [Address] = 'TCEQ, Contact Name, Mail Code, P.O. Box 13087',
           City = 'Austin',
           [State] = 'Texas',
           ZipCode = '78711-3087'

Select     *
from       #Sources_Dist

insert     into       Sources    (Organization,
                                 SourceDescription,
                                 SourceLink,
                                 ContactName,
                                 Phone,
                                 Email,
                                 [Address],
                                 City,
                                 [State],
                                 ZipCode)

           Select                Organization,
                                 SourceDescription,
```

```
                              SourceLink,
                              ContactName,
                              Phone,
                              Email,
                              [Address],
                              City,
                              [State],
                              ZipCode
          From      #Sources_Dist

Select    *
from      Sources

Drop Table #Sources_Dist

rollback tran
--commit tran


--**************VARIABLES
begin tran

Select    *
from      Variables

Select    [Storet Code] as variablecode, [Long Description] as
variablename
Into      #Storet_Codes
From      dbo.TRACS_Parameter

Alter Table #Storet_Codes ADD [SampleMedium] char(50) Null;
go
Alter Table #Storet_Codes ADD [ValueType] char(50) Null;
go
Alter Table #Storet_Codes ADD [GeneralCategory] char(50) Null;
go

update    #Storet_Codes
set       SampleMedium = 'Surface Water',
          ValueType = 'Field Observation',
          GeneralCategory = 'Water Quality'

Select    *
from      #Storet_Codes

insert    into      Variables (VariableCode,
                    VariableName,
                    SampleMedium,
                    ValueType,
                    GeneralCategory)
          Select    VariableCode,
                    VariableName,
                    SampleMedium,
```

```
                    ValueType,
                    GeneralCategory
          From      #Storet_Codes

Select    *
from      Variables

Drop Table #Storet_Codes

rollback tran
--commit tran

--***************Categories - No Match in SWQM

--***************Values
begin tran

Select    *
from      [values]

SELECT    TRACS_Results.Tag_ID,
          TRACS_Results.Gtlt as Censorcode,
          TRACS_Results.[Value],
          TRACS_Event.EndDate,
          TRACS_Event.EndTime,
          TRACS_Event.EndDepth as Offsetvalue,
          TRACS_Results.StoretCode,
          TRACS_Event.Source1,
          TRACS_Event.Source2,
          TRACS_Event.Program,
          TRACS_Event.StationID,
          TRACS_Event.Comment,
          TRACS_Event.Category,
          TRACS_Event.Type
Into      dbo.ValuesTemp
FROM      TRACS_Results INNER JOIN TRACS_Event
ON        TRACS_Results.Tag_ID = TRACS_Event.Tag_ID

Alter Table dbo.ValuesTemp ADD [GroupID] int Null;
go
Alter Table dbo.ValuesTemp ADD [Datetime] datetime Null;
go
Alter Table dbo.ValuesTemp ADD [VariableID] Int Null;
go
Alter Table dbo.ValuesTemp ADD [SourceID] int Null;
go
Alter Table dbo.ValuesTemp ADD [SiteID] int Null;
go
Alter Table dbo.ValuesTemp ADD [QualifierID] int Null;
go
Alter Table dbo.ValuesTemp ADD [SampleID] int Null;
go
Alter Table dbo.ValuesTemp ADD [OffsetTypeID] int Null;
```

```
go
Alter Table dbo.ValuesTemp ADD [Cattype] varchar(50) Null;
go
Alter Table dbo.ValuesTemp ADD [Sources] varchar(50) Null;
go
Alter Table dbo.ValuesTemp ADD [UTCOffset] int Null;
go
Alter Table dbo.ValuesTemp ADD [QualityControlLevel] int Null;
go
Alter Table dbo.ValuesTemp ADD [UTCDateTime] datetime Null;
go

Update    dbo.ValuesTemp
Set       GroupID = GroupDescriptions.GroupID
FROM      dbo.ValuesTemp INNER JOIN GroupDescriptions
ON        dbo.ValuesTemp.Tag_ID = GroupDescriptions.GroupDescription

Update    dbo.ValuesTemp
Set       censorcode = 'lt'
where     censorcode = '<'

Update    dbo.ValuesTemp
Set       censorcode = 'gt'
where     censorcode = '>'

Update    dbo.Valuestemp
set       [datetime] = enddate

Update    dbo.Valuestemp
set       [datetime] = [datetime] + endtime

Update    dbo.Valuestemp
set       [UTCdatetime] = dateadd(hh, 6, [datetime])

update    dbo.Valuestemp
set       [VariableID] = Variables.VariableID
FROM      dbo.Valuestemp INNER JOIN Variables
ON        dbo.Valuestemp.StoretCode = Variables.VariableCode

Update    dbo.Valuestemp
set       [SiteID] = Sites.SiteID
FROM      dbo.ValuesTemp INNER JOIN Sites
ON        dbo.ValuesTemp.StationID = Sites.SiteCode

Update    dbo.Valuestemp
set       [QualifierID] = Qualifiers.QualifierID
FROM      dbo.ValuesTemp INNER JOIN Qualifiers
ON        dbo.ValuesTemp.Comment = Qualifiers.QualifierDescription

Update    dbo.Valuestemp
set       OffsetTypeid = 1

Update    dbo.Valuestemp
```

```
set       cattype = category + '_' + [type]

update    dbo.Valuestemp
set       SampleID = Samples.SampleID
FROM      dbo.ValuesTemp INNER JOIN Samples
ON        dbo.ValuesTemp.cattype = samples.sampletype

update    dbo.Valuestemp
set       Sources = 'TCEQ Sources = ' + Source1 + '; ' + Source2 + '; '
+ Program

update    dbo.Valuestemp
set       SourceID = Sources.SourceID
FROM      dbo.ValuesTemp INNER JOIN Sources
ON        dbo.ValuesTemp.sources = sources.Organization

update    dbo.Valuestemp
set       UTCOffset = -6

update    dbo.Valuestemp
set       QualityControlLevel = 2

Select    *
from      dbo.ValuesTemp

insert into dbo.[Values]
     (
     [Value],
     [DateTime],
     UTCOffset,
     [UTCDateTime],
     SiteID,
     VariableID,
     OffsetValue,
     OffsetTypeID,
     CensorCode,
     QualifierID,
     SourceID,
     SampleID,
     QualityControlLevel,
     tag_id
     )
Select
     [Value],
     [DateTime],
     UTCOffset,
     [UTCDateTime],
     SiteID,
     VariableID,
     OffsetValue,
     OffsetTypeID,
     CensorCode,
     QualifierID,
```

```
     SourceID,
     SampleID,
     QualityControlLevel,
     tag_id
From      dbo.Valuestemp

Select    *
from      dbo.[Values]

Drop Table dbo.ValuesTemp

rollback tran
--commit tran

--***************Groups
begin tran

Select    *
from      Groups

SELECT
          [Values].ValueID,
          GroupDescriptions.GroupID
INTO      #GroupsTemp
FROM      [Values] INNER JOIN GroupDescriptions
ON        [Values].Tag_ID = GroupDescriptions.GroupDescription

SET IDENTITY_INSERT dbo.groups ON
GO

insert into groups
     (
     ValueID,
     GroupID
     )
Select
     ValueID,
     GroupID
From #groupstemp

Select    *
from      Groups

SET IDENTITY_INSERT dbo.groups OFF
GO

rollback tran
--commit tran

--***************SeriesCatalog - The following script to create the
--SeriesCatalog was created by Jeff Horsburgh at Utah State University
(jeffh@cc.usu.edu)
```

```
begin tran

--Clear out the entire SeriesCatalog Table
DELETE FROM  [SeriesCatalog];

--Recreate the records in the SeriesCatalog Table
INSERT INTO [SeriesCatalog]
SELECT      dbo.Sites.SiteID,
            dbo.Sites.SiteCode,
            dbo.Sites.SiteName,
            dbo.Variables.VariableID,
            dbo.Variables.VariableCode,
            dbo.Variables.VariableName,
        dbo.Variables.VariableUnitsID,
            dbo.Units.UnitsName AS VariableUnitsName,
            dbo.Variables.SampleMedium,
            dbo.Variables.ValueType,
        dbo.Variables.TimeSupport,
            dbo.Variables.TimeUnitsID,
            dbo.Units.UnitsName,
            dbo.Variables.DataType,
            dbo.Variables.GeneralCategory,
        MIN(dbo.[Values].DateTime) AS BeginDateTime,
            MAX(dbo.[Values].DateTime) AS EndDateTime,
            dbo.[Values].UTCOffset,
            COUNT(dbo.[Values].Value) AS ValueCount

FROM        dbo.Sites INNER JOIN
                  dbo.[Values] ON dbo.Sites.SiteID =
dbo.[Values].SiteID INNER JOIN
                  dbo.Variables ON dbo.[Values].VariableID =
dbo.Variables.VariableID INNER JOIN
                  dbo.Units ON dbo.Variables.VariableUnitsID =
dbo.Units.UnitsID

GROUP BY    dbo.Sites.SiteID,
                  dbo.Sites.SiteCode,
                  dbo.Sites.SiteName,
                  dbo.Variables.VariableID,
                  dbo.Variables.VariableCode,
                  dbo.Variables.VariableName,
                  dbo.Variables.VariableUnitsID,
                  dbo.Units.UnitsName,
                  dbo.Variables.SampleMedium,
                  dbo.Variables.ValueType,
                  dbo.Variables.GeneralCategory,
                  dbo.[Values].UTCOffset,
                  dbo.Variables.TimeSupport,
                  dbo.Variables.TimeUnitsID,
                  dbo.Variables.DataType

ORDER BY    dbo.Sites.SiteID,
                  dbo.Variables.VariableID,
```

```
                  dbo.Variables.VariableUnitsID

--Now Update the TimeUnitsName Field since I can only have a single
join on the Units table
UPDATE    SeriesCatalog
SET       TimeUnitsName = Units.UnitsName
FROM      SeriesCatalog INNER JOIN
          Units ON SeriesCatalog.TimeUnitsID = Units.UnitsID

rollback tran
--commit tran

--Template
/*
--***************
begin tran

Select    *
from

insert    into       ()
          Select
          From      dbo.TRACS_

Select    *
from

rollback tran
--commit tran
*/
```

References

Alameda, J. C. (2006), Web Services, in *Hydroinformatics : data integrative approaches in computation, analysis, and modeling* edited, pp. 163-175, Taylor & Francis, Boca Raton.

Booth, D., et al. (2004), Web Services Architecture, *WC3 Working Group Note*, 98.

CUAHSI (2007) Hydrologic Information System, <http://www.cuahsi.org/his.html> (March 10, 2007)

Folk, M. J. (2006), Data Formats, in *Hydroinformatics : data integrative approaches in computation, analysis, and modeling* edited, pp. 117-134, Taylor & Francis, Boca Raton.

Goodall, J. L. (2005), A Geotemporal Framework for Hydrologic Analysis, 192 pp, The University of Texas at Austin, Austin. <http://www.crwr.utexas.edu/reports/pdf/2005/rtp05-08.pdf>

Iowa State University (2007) Hydrologic Cycle, <http://www.buffer.forestry.iastate.edu/Photogallery/illustrations/Images/Hydrologic-Cycle.jpg> (March 9,2007)

Kumar, P., et al. (2006), *Hydroinformatics : data integrative approaches in computation, analysis, and modeling* 534 pp., Taylor & Francis, Boca Raton.

Maidment, D. R. (2005), Hydrologic Information System Status Report, 214 pp, CUAHSI. <http://www.cuahsi.org/his/docs/HISStatusSept15.pdf>

NOAA National Data Center (2007) Climate Data Online, <http://cdo.ncdc.noaa.gov/CDO/cdo> (3/12/07)

PRISM Group, O. S. U. (2006), July 2000 Precipitation, Inches, in *ArcGIS*, edited, Oregon State University.

Ruddell, B. L., and P. Kumar (2006), Hydrologic Data Models, in *Hydroinformatics : data integrative approaches in computation, analysis, and modeling* edited, pp. 61-80, Taylor & Francis, Boca Raton.

SAHRA (2006), SAHRA Strategic Plan. <http://www.sahra.arizona.edu/publications/StrategicPlan/SAHRA%20Strategic%20Plan.pdf>

Soh, L.-K., et al. (2006), A Task-Based Approach to User Interface Design for a Web-Based Hydrologic Information Systems, *Transactions in GIS*, *10*(3), 417-449.

Tarboton, D. G., et al. (2006), CUAHSI Community Observations Data Model Working Design Specifications Document – Version 4, 42 pp, CUAHSI. <http://www.cuahsi.org/his/docs/ODM4-20061026.pdf >

Tarboton, D. G., et al. (2007), CUAHSI Community Observations Data Model (ODM) Version 1.0 Design Specifications, 42 pp, CUAHSI. <http://www.cuahsi.org/his/docs/ODM1.pdf>

TCEQ (2006), Surface Water Quality Monitoring Data Management Reference Guide, 116 pp. <http://www.tceq.state.tx.us/assets/public/compliance/monops/water/wdma/dmrg/2005/2005dmrg_complete.pdf >

TCEQ (2007a) SWQM Event File Structure , <ftp://www.tceq.state.tx.us/pub/WaterResourceManagement/WaterQuality/DataCollection/CleanRivers/public/event_struct.txt> (February 2, 2007)

TCEQ (2007b) SWQM Results File Structure, <ftp://www.tceq.state.tx.us/pub/WaterResourceManagement/WaterQuality/DataCollection/CleanRivers/public/result_struct.txt> (February 2, 2007)

TCEQ (2007c) SWQM Stations File Structure, <ftp://www.tceq.state.tx.us/pub/WaterResourceManagement/WaterQuality/DataCollection/CleanRivers/public/stnsmeta.txt> (February 2, 2007)

TCEQ (2007d) SWQM Parameters File Structure, <ftp://ftp.tceq.state.tx.us/pub/WaterResourceManagement/WaterQuality/DataCollection/CleanRivers/public/parm_meta.txt> (February 2, 2007)

Texas Commission on Environmental Quality (2006), Surface Water Quality Monitoring Data Management Reference Guide, 116 pp. <http://www.tceq.state.tx.us/assets/public/compliance/monops/water/wdma/dmrg/2005/2005dmrg_complete.pdf >

Trujillo, D. N. (2006), Texas Surface Water Quality Monitoring Information System, The University of Texas at Austin. <https://webspace.utexas.edu/dt385/www/Term%20paper.htm>

USEPA (2007), What's New in STORET, 2 pp. <http://www.epa.gov/storet/wqx_products/Whats_new_in_STORET.pdf>

USGS (2000), The National Hydrography Dataset: Concepts and Contents, 75 pp. <http://nhd.usgs.gov/chapter1/chp1_data_users_guide.pdf>

USGS (2007) National Hydrography Dataset Home Page, <http://nhd.usgs.gov/> (December 2, 2007)

Whiteaker, T., et al. (2006), CUAHSI Web Services Workbook, 91 pp, CUAHSI. <http://www.cuahsi.org/his/docs/HIS-workbook-20061130.pdf>

Vita

Tyler Landis Jantzen was born to Jonathan Landis Jantzen and Mary Beth Landis Jantzen on November 13, 1981 in Albuquerque, New Mexico. He received his primary education from schools in Albuquerque; Jakarta, Indonesia; Longmont, Colorado and Tucson, Arizona where he graduated from University High School in 2000. Upon graduation he attended Gonzaga University in Spokane, Washington where he graduated Summa Cum Laude with a Bachelors of Science degree in Civil Engineering in 2005. While at Gonzaga he was a member of the American Society of Civil Engineers Student Chapter, Tau Beta Pi honor society, and Engineers Without Borders where he served as club Vice President. During his undergraduate work he also worked in engineering and construction with Granite Construction and the Spokane County Department of Public Works. He also worked as a CAD Technician with the Gonzaga University Campus Architect. Upon graduation from Gonzaga he attended the University of Texas at Austin, where he intends to obtain a Masters of Science in Environmental and Water Resources Engineering degree in May, 2007. After graduation he plans to work as a Water Resources Engineer with the company CH2M Hill in Bellevue, Washington.

Permanent address: 5434 East Holmes Street, Tucson, AZ 85711

This thesis was typed by the author.